

**UNIVERZITA KARLOVA V PRAZE
FILOZOFICKÁ FAKULTA
ÚSTAV INFORMAČNÍCH STUDIÍ A KNIHOVNICTVÍ**

Informační studia a knihovnictví – Informační věda

Eva B r a t k o v á

**Digitální knihovny s volným přístupem v oblasti vědy a výzkumu
a identifikace a metadatový popis jejich objektů**

**Digital libraries with open access in science and research area
and identification and metadata description of their objects**

Disertační práce

Školitel - doc. PhDr. Rudolf Vlasák

2009-03-30

P r o h l á š e n í

Prohlašuji, že jsem disertační práci vykonala samostatně s využitím uvedených pramenů a literatury.

V Praze, 30. března 2009

.....

podpis

Abstrakt (CZ)

Disertační práce je věnována analýze volně dostupných digitálních knihoven ve vědě a výzkumu a v praktické rovině předkládá české národní metadatové standardy připravené v rámci výzkumu a budování digitálních knihoven vysokoškolských kvalifikačních prací v České republice. Druhá kapitola představuje problematiku terminologie v oblasti digitálních knihoven a obecné modely (rámce) pro jejich budování, konkrétně klasický model „Kahn-Wilensky“ a referenční model Otevřeného archivního informačního systému OAIS se zaměřením na oblast institucionálních repozitářů. Třetí kapitola zahrnuje otázky vývoje a současného stavu otevřeného přístupu (angl. „open access“) k informacím v oblasti vědy a výzkumu, jeho hlavních zdrojů včetně iniciativ na jeho podporu. Čtvrtá, rozsáhlá kapitola představuje vybrané systémy volně dostupných elektronických archivů nebo digitálních knihoven ve vědě a výzkumu. Zahrnuty jsou vybrané systémy a služby z oboru počítačová věda, ekonomika, oboru informační a knihovní věda, sociálních věd, specifický britský systém pro vědu a výzkum (Intute), digitální knihovny s citačními službami a digitální knihovny VŠKP. Pátá kapitola představuje výsledek praktické výzkumné činnosti – české národní metadatové standardy pro popis a komunikaci vysokoškolských kvalifikačních prací (EVSKP-MS, PersCZ a CorpCZ) s připojenou problematikou k trvalým identifikátorům informačních entit.

Abstract (EN)

Thesis is devoted to the analysis of open accessible digital libraries in science and research and, at practical level, thesis puts Czech national metadata sets prepared within research and ETD digital libraries development in the Czech Republic. The second chapter presents terminology issues in digital libraries area and general models of digital libraries - „Kahn-Wilensky“ architecture and Reference Model Open Archival Information System (OAIS) with special regard to institutional repositories. The third chapter contains the survey of development and current state of open access to information in science and research and its main resources including initiatives of its support. The fourth chapter presents selected systems of open access electronic archives or digital libraries in science and research. Included are selected systems and services in computer science, economics, library and information science, social science, the specific British system of internet resources Intute, digital libraries with citation services and ETDs digital libraries. The fifth chapter presents the outputs of research – Czech national metadata sets for ETDs description and communication with added issues of persistent identifiers of information entities.

P o d ě k o v á n í

Autorka děkuje svému školiteli doc. PhDr. Rudolfovi V l a s á k o v i za veškerou pomoc poskytnutou v průběhu přípravy disertační práce.

OBSAH

1	PŘEDMĚT, CÍLE A METODY PRÁCE	9
2	VYBRANÉ OBECNÉ OTÁZKY DIGITÁLNÍCH KNIHOVEN	11
2.1	ÚVODEM	11
2.2	ČASOVÝ RÁMEC VÝVOJE DIGITÁLNÍCH KNIHOVEN	11
2.3	PRACOVNÍ VYMEZENÍ HLAVNÍCH TERMÍNŮ	12
2.4	PROBLEMATIKA DEFINICE DIGITÁLNÍCH KNIHOVEN A JEJICH FUNKCÍ	15
2.5	PROBLEMATIKA TYPOLOGIE DIGITÁLNÍCH OBJEKTŮ (ELEKTRONICKÝCH ZDROJŮ).....	17
2.5.1	<i>Typologie elektronických zpráv v rámci internetové normy MIME.....</i>	<i>18</i>
2.5.2	<i>Typologie informačních zdrojů standardu Dublin Core</i>	<i>20</i>
2.6	ARCHITEKTURA DIGITÁLNÍ KNIHOVNY PODLE AMERICKÉHO MODELU „KAHN-WILENSKY“	24
2.6.1	<i>Identifikátory a systémy pro jejich správu a směrování.....</i>	<i>25</i>
2.6.2	<i>Digitální objekt.....</i>	<i>28</i>
2.6.3	<i>Skupina (agregát) digitálních objektů.....</i>	<i>29</i>
2.6.4	<i>Digitální repozitáře (skladiště, úložiště)</i>	<i>30</i>
2.7	REFERENČNÍ MODEL OTEVŘENÉHO ARCHIVNÍHO INFORMAČNÍHO SYSTÉMU (RM OAIS)	30
2.7.1	<i>Vznik mezinárodní normy RM OAIS</i>	<i>31</i>
2.7.2	<i>Archiv typu OAIS jako ústřední pojem a jeho hlavní odpovědnosti (funkce)...</i>	<i>32</i>
2.7.3	<i>Externí entity archivu OAIS a jejich odpovědnosti (funkce)</i>	<i>33</i>
2.7.4	<i>Informace, informační objekty a balíčky informací a jejich pojetí v modelu OAIS</i>	<i>34</i>
2.7.5	<i>Funkční model archivu typu OAIS (základní úroveň)</i>	<i>38</i>
2.7.6	<i>Referenční model OAIS a praxe</i>	<i>41</i>
2.7.6.1	<i>Modely využití RM OAIS v oblasti institucionálních repozitářů</i>	<i>42</i>
3	OTEVŘENÝ PŘÍSTUP	44
3.1	DEFINICE OTEVŘENÉHO PŘÍSTUPU	44
3.2	VÝZNAMNÉ UDÁLOSTI PŘED NÁSTUPEM HNUTÍ OTEVŘENÉHO PŘÍSTUPU (DO ROKU 1991)	45
3.3	HLAVNÍ ZDROJE S OTEVŘENÝM PŘÍSTUPEM K INFORMACÍM (VÝVOJ VOLNÉHO PŘÍSTUPU PO ROCE 1991).....	46
3.3.1	<i>Mezinárodní elektronické archivy a digitální knihovny.....</i>	<i>46</i>
3.3.2	<i>Institucionální repozitáře</i>	<i>49</i>
3.3.3	<i>Registry elektronických archivů</i>	<i>51</i>
3.3.4	<i>Volně dostupné online časopisy</i>	<i>52</i>
3.3.5	<i>Volně dostupné vědecké a jiné dokumenty na WWW</i>	<i>53</i>
3.4	PODPORA OTEVŘENÉHO PŘÍSTUPU K INFORMACÍM A JEHO RŮST	53
3.5	INICIATIVA OTEVŘENÝCH ARCHIVŮ OAI A SYSTÉM ARC	56
4	VYBRANÉ SYSTÉMY DIGITÁLNÍCH ARCHIVŮ A DIGITÁLNÍCH KNIHOVEN S OTEVŘENÝM PŘÍSTUPEM VE VĚDĚ A VÝZKUMU (ANALÝZA) 60	
4.1	ELEKTRONICKÉ ARCHIVY A DIGITÁLNÍ KNIHOVNY Z OBORU POČÍTAČOVÉ VĚDY	61
4.1.1	<i>Archiv elektronických tisků arXiv.org.....</i>	<i>61</i>
4.1.2	<i>NCSTRL (Networked Computer Science Technical Reference Library)</i>	<i>66</i>
4.1.3	<i>CoRR (Computing Research Repository).....</i>	<i>69</i>
4.1.4	<i>Institucionální repozitář elektronických tisků ECS.....</i>	<i>71</i>

4.1.4.1	Úvodní charakteristika repozitáře	71
4.1.4.2	Obsahové vymezení repozitáře ECS	71
4.1.4.3	Typologie dokumentů	72
4.1.4.4	Vyhledávání informací	72
4.1.5	<i>Digitální knihovna Nového Zélandu NZDL (část počítačové vědy)</i>	74
4.1.5.1	Úvodní charakteristika systému	74
4.1.5.2	Model digitální knihovny NZDL/CSTR (vstupy a jejich zpracování)	75
4.1.5.3	Vyhledávání informací z digitální knihovny NZDL	76
4.2	SYSTÉM DIGITÁLNÍ KNIHOVNY PRO OBLAST EKONOMIKY RePEc	77
4.2.1	Úvodní charakteristika systému	77
4.2.2	Účastníci systému RePEc	79
4.2.3	Typy zahrnovaných dokumentů a nástroje pro zpracování metadat	79
4.2.4	Služby systému RePEc	80
4.2.4.1	Registrace autorů v rámci Autorské služby RePEc	81
4.2.4.2	Registrace institucí v rámci služby EDIRC	83
4.2.4.3	Služba IDEAS a registrační databáze systému RePEc	83
4.2.4.4	Mnichovský personální archiv RePEc (MPRA)	85
4.2.4.5	EconPapers	86
4.2.4.6	Služba NEP	87
4.2.4.7	LogEc	87
4.2.4.8	CitEc	89
4.2.5	Závěr	90
4.3	ELEKTRONICKÉ ARCHIVY A DIGITÁLNÍ KNIHOVNY OBORU KNIHOVNÍ A INFORMAČNÍ VĚDY	91
4.3.1	<i>Archiv elektronických tisků E-LIS</i>	91
4.3.1.1	Organizační, programové a technické zabezpečení archivu	91
4.3.1.2	Ukládání dokumentů do archivu E-LIS	93
4.3.1.2.1	Přijímané typy a formáty dokumentů	93
4.3.1.2.2	Aktuální stav repozitáře archivu	94
4.3.1.3	Metadata pro zpracování ukládaných dokumentů	95
4.3.1.4	Oborová klasifikace JITA	96
4.3.1.5	Ukládání informací	97
4.3.1.6	Autorská práva	99
4.3.1.7	Výstupy a služby archivu E-LIS	100
4.3.1.7.1	Online vyhledávání a zobrazování informací	100
4.3.1.7.2	Další služby archivu	104
4.3.1.8	Závěr	105
4.3.2	<i>Archiv elektronických tisků pro vědy o informacích a komunikaci „@rchiveSIC“</i>	105
4.3.3	<i>Archiv elektronických tisků pro informační vědu a technologie DLIST</i>	110
4.3.4	<i>Elektronická knihovna a alternativní bibliografická služba pro informační vědu DoIS</i>	115
4.3.5	<i>Závěry</i>	118
4.4	ELEKTRONICKÝ ARCHIV COGPRINTS	120
4.5	PŘEDMĚTOVÉ BRÁNY SYSTÉMU INTUTE (DŘÍVE RDN A ROADS)	121
4.5.1	<i>Počátky - předmětové brány v systému ROADS</i>	122
4.5.2	<i>Síť předmětových bran RDN</i>	123
4.5.3	<i>Systém předmětových bran Intute</i>	125
4.6	VOLNĚ DOSTUPNÉ DIGITÁLNÍ KNIHOVNY S CITAČNÍMI SLUŽBAMI	130
4.6.1	Úvod	130

4.6.2	<i>Citační indexace a citační služby</i>	131
4.6.3	<i>Vybrané volně dostupné systémy s citačními službami</i>	131
4.6.3.1	Google Scholar	131
4.6.3.2	CiteSeer	133
4.6.3.3	Citebase Search	137
4.6.4	<i>Závěr</i>	140
4.7	DIGITÁLNÍ KNIHOVNY VYSOKOŠKOLSKÝCH KVALIFIKAČNÍCH PRACÍ	142
4.7.1	<i>Úvod</i>	142
4.7.2	<i>Zpřístupňování elektronických disertací v zahraničí</i>	143
4.7.2.1	Mezinárodní federace a systém ND LTD	143
4.7.2.2	Mezinárodní systém e-disertací „CyberThèses“	147
4.7.2.3	Australasijský systém pro digitální disertace (ADT)	149
4.7.2.4	Národní německý systém „DissOnline.de“	151
4.7.2.5	Situace s rozšiřováním vysokoškolských kvalifikačních prací v České republice	153
5	IDENTIFIKACE A POPIS DIGITÁLNÍCH OBJEKTŮ V DIGITÁLNÍCH KNIHOVNÁCH	155
5.1	METADATA	155
5.2	METADATOVÉ STANDARDY PRO VŠKP	156
5.2.1	<i>Metadata pro elektronické disertace</i>	156
5.3	NÁVRH VÝVOJOVÉ VERZE NÁRODNÍHO STANDARDU EVSKP-MS	164
5.3.1	<i>Úvod</i>	164
5.3.2	<i>Specifikace metadat</i>	165
5.3.3	<i>Reprezentace metadatových prvků</i>	167
5.4	NÁRODNÍ METADATOVÝ STANDARD EVSKP-MS, v. 1.1. PRO POPIS VŠKP A STANDARDY SOUVISEJÍCÍ (AKTUÁLNÍ VERZE STA)	167
5.4.1	<i>Úvod</i>	167
5.4.2	<i>Popisné, technické a administrativní prvky EVSKP-MS</i>	169
5.4.2.1	Popisné prvky EVSKP-MS	170
5.4.2.2	Technické a administrativní prvky EVSKP-MS	172
5.4.3	<i>Popisné prvky standardů PersCZ a CorpCZ</i>	173
5.4.4	<i>Formát zápisu metadatových prvků a validace záznamu</i>	175
5.4.5	<i>Závěr</i>	182
5.5	VYBRANÉ TRVALÝCH IDENTIFIKAČNÍ PRVKY PRO DIGITÁLNÍ OBJEKTY (DOKUMENTY)	183
5.5.1	<i>Úvod</i>	183
5.5.2	<i>URI – obecný rámec trvalých identifikátorů</i>	184
5.5.3	<i>URN</i>	186
5.5.3.1	URN obecně	186
5.5.3.2	Jmenný prostor URN:NBN	188
5.5.4	<i>HDL (Handle)</i>	192
5.5.5	<i>DOI</i>	195
5.5.6	<i>PURL</i>	198
5.5.7	<i>AKR</i>	201
5.6	TRVALÁ IDENTIFIKACE FYZICKÝCH OSOB A KORPORACÍ V PROSTŘEDÍ DIGITÁLNÍCH ARCHIVŮ : AKTUÁLNÍ SMĚRY ŘEŠENÍ V ZAHRANIČNÍCH SYSTÉMECH	204
5.6.1	<i>Úvod do problematiky - sledovaný problém</i>	204
5.6.2	<i>Identifikace autoritních entit v knihovnické oblasti</i>	205
5.6.2.1	Plánované identifikátory ISADN a INSAN	206
5.6.2.2	Projekt „WorldCat Identities“ (OCLC) a služba LAF (LC a OCLC)	207

5.6.2.3	Projekt Virtuálního mezinárodního autoritního souboru (VIAF)	210
5.6.2.4	Trvalý identifikátor ARK pro záznamy autoritních entit	211
5.6.3	<i>Identifikátory v oblasti autorských ochranných svazů</i>	213
5.6.3.1	Identifikátor IPI (IPIN) v sektoru správy autorských práv	213
5.6.3.2	Mezinárodní identifikátor veřejných identit ISNI (projekt ISO normy)	213
5.6.4	<i>Trvalý identifikátor „ResearcherID“ (komerční sféra)</i>	215
5.6.5	<i>Identifikátory v oblasti digitálních repozitářů</i>	216
5.6.5.1	Trvalý Digitální identifikátor autora (Nizozemsko).....	216
5.6.5.2	Systém RePEc a jeho identifikátory autorů a korporací	219
5.6.6	<i>Závěr</i>	220
6	ZÁVĚRY PRÁCE	221
	POUŽITÉ INFORMAČNÍ ZDROJE.....	222
	SEZNAM POUŽITÝCH ZKRATEK.....	238
	PŘÍLOHY	242

1 Předmět, cíle a metody práce

Text předkládané disertační práce na téma *Digitální knihovny s volným přístupem v oblasti vědy a výzkumu a identifikace a metadatový popis jejich objektů* byl připravován postupně v celém období doktorského studia, od roku 2001. Autorka v něm ale navázala na předchozí publikované práce (před rokem 2001) vztahující se k tématu digitálních knihoven se zaměřením na oblast metadatového popisu a trvalé identifikace informačních objektů. Volba tématu úzce souvisela s výzkumnou a pedagogickou činností, kterou autorka zajišťuje na Ústavu informačních studií a knihovnictví (ÚISK) FF UK v Praze. Na přelomu tisíciletí byl do učebního programu zaveden předmět Digitální knihovny ve vědě a technice, bylo tedy žádoucí i formou doktorského studia doplnit teoreticky i prakticky znalosti ze stanovené oblasti. Výchozí projekt disertační práce byl zadán široce do oblasti digitálních knihoven ve vědě a technice, s postupem času se téma vymezovalo přesněji do oblasti digitálních knihoven ve vědě a výzkumu se zaměřením na fenomén druhé poloviny 20. století a počátku nového tisíciletí „otevřený přístup“ k informacím, v tomto případě k informacím v organizovaných fondech digitálních knihoven. V práci nepostihuje oblast digitálních knihoven komerční sféry, která je pro oblast vědy a výzkumu také významnou oblastí. Další vymezení práce je dáno zaměřením na oblast metadatového popisu a trvalé identifikace informačních digitálních objektů (entit), které autorka práce sleduje dlouhodobě – je specializovaná, kromě jiného na problematiku formátů pro bibliografická a jiná data. Je nutné zdůraznit, že práce není orientována na technické a technologické aspekty provozu digitálních knihoven, užívané softwarové aplikace, které používají zahraniční a domácí systémy digitálních knihoven jsou charakterizovány jen stručně. Okrajově je zmiňována také o složitá problematika autorských práv v oblasti digitálních knihoven (nebyla předmětem výzkumu).

Obsah celé disertační práce zahrnuje v podstatné míře již zveřejněné výsledky tvůrčí práce autorky (viz seznam použité literatury, část [BRATKOVÁ, 1999-2008]). Jde o publikace vzešlé především z výzkumu a vývoje. Autorka se podílela na práci na dvou oficiálních grantových úlohách: grantové úloze FRVŠ č. 2045/2003 (okruh E2) „E-DI“ (Digitální knihovna diplomových a disertačních prací UK v Praze) a na grantové úloze Centralizovaný projekt „Národní registr VŠKP a úložiště závěrečných prací se službou na odhalování plagiátů“, MŠMT C1/2008). Další publikace vzešly z vlastního bádání konaného v rámci doktorského studia a také pedagogické činnosti na ÚISK.

K praktickým zkušenostem, které byly pro disertační práci využity, patří také funkce editorky (za Českou republiku) v mezinárodním archivu E-LIS pro obor knihovní a informační vědy (<http://eprints.rclis.org/>), kterou autorka přijala a kterou vykoná od roku 2004. Šlo o cennou zkušenost, která přinesla nové informace ze „zákulísí“ volně dostupného vědecké elektronického archivu (v systému EPrints). Podstatné praktické zkušenosti byly získány v rámci působení v Odborné komisi pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací při Asociaci knihoven vysokých škol ČR (<http://www.evskp.cz/>), kterou pomohla v roce 2004 založit a již je odborným garantem a místopředsdkyní. Činnost komise je orientována na řešení palčivých otázek oblasti digitálních knihoven či digitálních archivů pro vysokoškolské kvalifikační práce. Jiné zkušenosti má autorka z provozu lokálního digitálního systém závěrečných prací ÚISK ISDP (<http://isdpl.alstanet.cz/>), který v roce 1999 založila a dodnes spravuje.

Hlavním cílem disertační práce bylo prozkoumat charakter a vývojové trendy zakládání a rozvíjení novodobých systémů digitálních knihoven a elektronických archivů s volným přístupem v oblasti vědy a výzkumu a obsahu i rozsahu jejich informačních fondů (jde v podstatě o sektor tzv. šedé literatury) se zvláštním – též praktickým - zaměřením na trvalou

identifikaci a popis jejich digitálních objektů. Podstatným výsledkem výzkumu autorky je podstatný podíl autorky na přípravě českých národních metadatových standardů pro popis a komunikaci informací o vysokoškolských kvalifikačních pracích a také informací o fyzických osobách a korporacích, tedy objektech (entitách), která s pracemi souvisí.

Disertační práce zahrnuje tři významné celky. První část má obecně teoretický charakter, autorka v ní prezentuje znalosti získané analýzou podstatné světové literatury a jiných zdrojů z vymezené oblasti a získané také z vyhodnocení vlastní praxe v této oblasti. Dílčí 2. kapitola osvětluje problematiku hlavních pojmů (digitální knihovna, elektronický archiv aj.), typologii systémů digitálních knihoven a typologii komunikovaných informačních objektů, metadatová schémata i obecné modely či rámce digitálních knihoven (klasický model Kahn-Wilensky a nový model OAIS (ISO 14721) určený pro dlouhodobou archivaci digitálních informací). 3. kapitola představuje fenomén „otevřeného přístupu“ k vědeckým informacím s využitím zahraničních výzkumů a publikací a také vlastních zkušeností.

Druhá část práce (rozsáhlá 4. kapitola) má systémově analytický charakter. Autorka v ní představuje výsledky dlouhodobě realizované analýzy vybraných systémů digitálních knihoven a elektronických archivů s volným přístupem v oblasti vědy a výzkumu (v globálním měřítku). Zvoleny byly systémy z oblasti počítačové vědy, informační a knihovní vědy, ekonomických studií; specifické místo je věnováno systémům pro zpřístupňování elektronických disertačních a jiných kvalifikačních prací (využito je analýzy z grantu FRVŠ č. 2045/2003, Digitální knihovna diplomových a disertačních prací UK v Praze) a také systémům citačních volně dostupných služeb). Sledovány jsou současné integrační tendence systémů digitálních knihoven šedé literatury. Analýzy vybraných systémů se zaměřují na jejich podstatné parametry (organizační a programové zajištění, obsah informačních fondů, datové a zejména metadatové formáty, trvalé identifikátory, rozhraní pro vyhledávání informací a informační služby a propojování systémů pomocí metadatových toků).

Třetí část práce (5. kapitola) má ryze praktický ráz. Autorka v ní představuje výsledky vlastní práce na prováděné v rámci výzkumu rozvojového projektu MŠMT C1/2008 „Národní registr VŠKP a úložiště závěrečných prací se službou na odhalování plagiátů“, kterého se účastnila za AMU (UK nebyla v projektu zastoupena), a to v oblasti přípravy národních standardů (vývojové a první verze) - výměnných metadatových formátů pro komunikaci vysokoškolských kvalifikačních prací a souvisejících objektů. Dílčí kapitoly zahrnují výsledky analýzy významných mezinárodních metadatových standardů (ADT, ETD-MS, xMetaDis aj.) a dále analytické komentáře k připraveným, schváleným a zveřejněným národním metadatovým standardům pro VŠ práce EVSKP-MS, pro popis fyzických osob PerCZ a pro popis korporací CorpCZ. Schválené standardy tvoří přílohu práce. Zahrnuta je i vývojová verze 0.1, která je od roku 2005 i nadále užívána pro komunikaci metadat v ČR. Připojena je související problematika trvalých identifikátorů, se kterými uvedené standardy počítají, jejichž zavedení se však v rámci ČR připravuje.

2 Vybrané obecné otázky digitálních knihoven

2.1 Úvodem

K významným zdrojům informací v oblasti vědy, výzkumu a vzdělávání patří dnes **digitální knihovny** (v širokém slova smyslu), popřípadě jejich dílčí typ **elektronické archivy** dostupné celosvětově v síti Internet. Jde o významné a podstatné organizované komponenty vytvářející se a stále se zdokonalující **globální informační infrastruktury (GII)**. K velké řadě z nich je přístup vázán na placenou licenci – jde zejména o významné digitální fondy článkových a monografických publikací renomovaných světových nakladatelů. Zatím menší, avšak narůstající množina těchto digitálních informačních zdrojů je již dostupná také volně a zdarma. Jde zejména o tzv. **archivy „elektronických tisků“** (centralizované předmětově profilované nebo institucionální repozitáře) provozované a naplňované přímo členy univerzitních či akademických pracovišť po celém světě. Lze konstatovat, že celý původní sektor tzv. **šedé literatury** (dříve definované jako literatury nepublikované, obtížně dostupné a identifikovatelné atd.) nastoupil ke svému novému perspektivnímu vývoji. Nikdy v minulosti nebyla šedá literatura tak snadno (technologicky), rychle a eventuálně i zdarma dostupná, jako tomu je v současné době. Velká skupina institucí (akademických a jiných) zahájila novodobé elektronické publikování, zpracování a ukládání (organizované v elektronických archivech) a zpřístupňování produktů vlastní vědecké a výzkumné činnosti.

Mnohé z volně dostupných (angl. „open access“) digitálních knihoven nebo elektronických archivů dnes poskytují velmi hodnotné a kvalitní informace v podobě úplných textů dokumentů, datových souborů nebo dalších typů informací z oblasti vědy, výzkumu a vysokoškolského vzdělávání. Díky novým informačním technologiím dochází k žádoucí **integraci** a dnes i k novému, **opakovanému užití** všech informačních zdrojů a také výzkumných a vzdělávacích aktivit ve prospěch uživatelů informací (viz například komplexní nizozemský systém informací pro vědu a výzkum NARCIS, <http://www.narcis.info/>). Řeší se další závažné otázky rozvíjení systémů digitálních knihoven, jako je například **dlouhodobá archivace a ochrana** informačních objektů, kterým doposud některé instituce nebo informační systémy nevěnovaly dostatečnou nebo vůbec žádnou pozornost.

Text disertační práce zahrnuje ve druhé kapitole **vybrané otázky k obecným otázkám digitálních knihoven** – terminologie, problematika typologie zdrojů, obecné modely (rámce) digitálních knihoven aj. Problematika metadat a trvalých identifikátorů je zařazena zcela výjimečně až v kap. 5.

2.2 Časový rámec vývoje digitálních knihoven

Budování a rozvoj **digitálních knihoven (DK)**, popř. **automatizovaných digitálních knihoven (ADK)**, probíhá zhruba od počátku 90. let 20. století. Intenzivnější rozvoj digitálních knihoven nastává po roce 1995, kdy se začala rychle uplatňovat služba WWW v rámci sítě Internet a vznikla možnost pohodlnějšího zpřístupňování jejich informací na dálku. (Pro domácí české uživatele, vědce, badatele a studenty se otevřely do té doby nevídané možnosti využívání významných zahraničních zdrojů, které se dnes jeví samozřejmostí). Desetiletý intenzivní rozvoj digitálních knihoven byl americkými odborníky v roce 2005 bilancován v prázdninovém čísle známého elektronického časopisu *D-Lib magazine* (<http://www.dlib.org/>, 2005, no. 7/8). Digitální knihovny se staly významným faktorem sociálního a ekonomického rozvoje společnosti, označované dnes také termínem **„informační společnost“**. Světovým „lídrem“ budování digitálních knihoven se staly Spojené státy americké, odkud se zaváděné modely, technologie a standardy postupně

rozšířily do dalších zemí. Probíhá čilá spolupráce při budování a rozvíjení systémů digitálních knihoven mezi různými subjekty i na mezinárodní úrovni. V rámci přístupu k informačním fondům digitálních knihoven přes počítačové sítě (Internet) se také někdy objevuje termín **digitální virtuální knihovna (DVK)**. Digitální virtuální knihovny a jejich informační zdroje (fondy, sbírky, kolekce apod.) představují významný typ **spravovaných a organizovaných informačních zdrojů** v současném internetu. Další významná etapa rozvoje digitálních knihoven nastala po roce 2000, kdy byla zahájena integrace a zejména dynamické propojování jednotlivých systémů digitálních knihoven a archivů v globálním měřítku. Vznikla řada iniciativ a virtuálních komunit, které tuto integraci pomáhají intenzivně rozvíjet včetně problematiky dlouhodobé archivace digitálních informací.

2.3 Pracovní vymezení hlavních termínů

V současné době existuje v celosvětovém rámci (v různých jazykových ekvivalentech) větší množství termínů k označování **systémů, které částečně nebo zcela automaticky zjišťují, získávají, zpracovávají, ukládají či archivují (zejména dlouhodobě), rozšiřují a zpřístupňují informace v elektronické formě (zejména textové povahy) přes internet**. Jednotlivé typy těchto systémů zahrnují různé funkce a procesy a mnohé z nich představují, jak bylo uvedeno výše, zároveň **nové formy elektronického publikování („elektronická vydavatelství“)** - např. digitální knihovny vydavatelů, archivy elektronických tisků, tj. vědeckých preprintů či postprintů aj. Většina z existujících termínů pro označení takovýchto digitálních systémů zahrnuje slovo (přídavné jméno) „**digitální**“, které znamená, že **předmětem jejich činnosti** jsou informace v digitální formě (tj. digitální objekty, resp. digitální zdroje informací v definovaných formátech, reprezentovaných „kombinacemi jedniček a nul“ či „proudem bitů“). Čistě český ekvivalent „**číslicový**“ se ve slovním spojení „**číslicová knihovna**“ v české praxi neujal (na rozdíl od některých jiných jazyků, například francouzštiny, kde se běžně užívá termínu „bibliothèque numérique“, i když i ten může být v heslářích odkazován na termín „bibliothèque virtuelle“). Za synonymum termínu „**digitální**“ může být považován termín „**elektronický**“ (nejde ovšem o čistou synonymii). Další slova (podstatná jména) v níže uvedených slovních spojeních jsou převzata ze slovníků tradičních systémů pro zpřístupňování informací (knihovna, archiv, muzeum aj.). To však neznamená, že by dnes budované digitální knihovny, elektronické archivy apod. byly nutně budovány na institucionálním principu (ten může a nemusí stát v pozadí). Situace je složitější. Pro počítačové odborníky například institucionální hledisko hraje minimální úlohu, jestli ji vůbec hraje. Příznačné pro současné období také je, že digitální či digitalizované informace připravuje a poskytuje dnes daleko širší spektrum institucí než dříve, a to i takových, které se této činnosti v tradičních podmínkách buď věnovaly málo nebo vůbec ne. Na pořadu dne je jejich efektivní integrace v podmínkách počítačových sítí. V zájmu uživatelů vznikají nadstavbové systémy a služby pracující nad „primárními“ digitálními knihovnami. Některá slovní spojení zahrnují také slovo „**virtuální**“, které znamená především způsob (cestu) zpřístupňování či poskytování digitálních (popř. digitalizovaných) informací prostřednictvím počítačových sítí (uživatel se „pohybuje“ v dané knihovně, sbírce atd. prostřednictvím obrazovky svého klientského počítače).

Ve světové i domácí české praxi, ve webových prezentacích i v literatuře můžeme najít následující frekventované odborné termíny z oblasti digitálních knihoven (některé z nich nemají dosud jednotné výklady):

- **Digitální knihovna** (angl. Digital Library, něm. Digitale Bibliothek, fr. Bibliothèque Numérique) - základní a zatím nejpoužívanější termín pro informační systémy, zpravidla integrovaného typu, které zabezpečují zjišťování, získávání, ukládání, zpracování, ukládání (v lepším případě i dlouhodobou archivaci) a zejména rozšiřování

a zpřístupňování informací v elektronické formě (elektronických informačních zdrojů); jejich podstatnou komponentou je repozitář neboli digitální sklad; digitální knihovny se vyznačují významnou škálou informačních služeb, které se stále rozšiřují; slovo „digitální“ zdůrazňuje povahu zpracovávaných dat a digitální knihovny zpřístupňují informace především prostřednictvím počítačových sítí; v řadě oborových tezaurů nebo řízených heslářů je tento pojem důležitým deskriptorem („Digital libraries“), je tomu tak například v tezauru databáze LISTA (EBSCO) (termíny „Electronic libraries“ a „Virtual libraries“ jsou v tomto tezauru zásadně odkazovány jako nedeskripty), dále v tezauru databáze LISA (CSA) (termín „Virtual libraries“ je zde ale už svébytným asociovaným deskriptorem a termín „Electronic libraries“ není uveden vůbec); další informace k problematice definice pojmu „digitální knihovna“ a také popis modelů digitálních knihoven je uveden v části - viz části 2.4-2.7

- **Elektronická knihovna** (angl. Electronic Library) - v podstatě jde o synonymum či ekvivalent termínu „digitální knihovna“ (jak sděluje též terminologická databáze TDKIV [KTD, 2003]); slovo „elektronická“ však zdůrazňuje aspekt využívání technických prostředků pro elektronickou komunikaci dat; tento termín se objevuje spíše v projektech a v literatuře britské provenience (například název známého britského programu 90. let pro budování digitálních knihoven zněl „E-Lib“), ale i některé americké systémy upřednostňují tento termín před termínem „digitální knihovna“ - například v tezauru databáze ERIC je deskriptorem termín „Electronic Libraries“ a termín „Digital Libraries“ i termín „Virtual libraries“ jsou jen odkazovány jako nedeskripty; česká databáze TDKIV [KTD, 2003] považuje termín „Elektronická knihovna“, stejně jako termín „Virtuální knihovna“ za ekvivalent upřednostňovaného termínu „Digitální knihovna“
- **Automatizovaná digitální knihovna** (angl. Automated digital library) - termín pro systémy digitálních knihoven, které mají všechny procesy a operace zajišťované počítačovými programy (v optimální míře bez zásahu člověka, kromě správce počítačového systému); tento termín ve svém článku použil a osvětlil americký expert na digitální knihovny počítačový vědec prof. W. Y. Arms [ARMS, 2000a]
- **Virtuální knihovna** (angl. Virtual Library) - obecný termín pro systémy digitálních či elektronických knihoven jakéhokoli typu, které jsou dostupné přes počítačové sítě; slovo „virtuální“ zdůrazňuje aspekt přístupu k systému na dálku; jak bylo uvedeno výše; termín „virtuální knihovna“ nemá zatím jednotný výklad; lze připojit, že v tezauru databáze LISA je tento termín na úrovni deskriptoru a termín „Electronic libraries“ je z něho odkazován jako nedeskriptor; v české národní databázi autoritních záznamů AUT (NK ČR) existuje předmětové heslo „digitální knihovny“ a také „virtuální knihovny“ (obě ve vztahu souvisejících termínů), zatímco termín „elektronické knihovny“ není prozatím uveden vůbec - ani na úrovni odkazu
- **Digitální virtuální knihovna** (angl. Digital Virtual Library) - termín pro systémy digitálních knihoven dostupných na dálku přes počítačové sítě (tento termín se objevil v některých dřívějších projektech, například v Německu)
- **Kybernetická knihovna** (angl. Cyber Library) - málo frekventovaný termín pro označení zcela automatizované digitální knihovny, která plní své funkce bez zásahu člověka v prostředí počítačových sítí; v praxi světového webu se lze setkat s některými aplikacemi takto pojmenovanými, ovšem při bližším prozkoumání se ukáže, že jde o běžné systémy digitálních či virtuálních knihoven; v dané chvíli by mohlo jít eventuálně i o futuristický termín pro označení budoucích systémů digitálních knihoven
- **Digitální sbírka, popř. digitální fond** (angl. Digital Collection) - termín pro označení buď podstatné komponenty digitální knihovny, nebo termín pro označení jednoduché formy

digitální knihovny, která nemusí zajišťovat všechny její charakteristické funkce (například může chybět interaktivní vyhledávání)

- ***Elektronický [Digitální] archiv*** (angl. Electronic [Digital] Archive) - obecný termín pro systémy, v jejichž rámci jsou ukládány digitální nebo digitalizované materiály (textové, obrazové i jiné povahy); zajišťována je prioritně archivní funkce, řeší se zejména dlouhodobá archivace; v současnosti jsou běžně budovány vydavateli publikací, tradičními archivy, které digitalizují archiválie, národními knihovnami, které digitalizují vlastní knihovní fondy, zejména historického charakteru, vědeckými komunitami apod.
- ***Archiv elektronických tisků*** (angl. E-Print Archive) - specifický frekventovaný termín pro systémy fungující v oblasti vědy; jsou určené k online ukládání a vyhledávání tzv. „elektronických tisků“ (tj. preprintů nebo postprintů vědeckých článků v různé fázi jejich tvorby a posuzování) včetně metadat (na bázi auto-archivace samotnými autory) v prostředí počítačových sítí; jsou zároveň jednou z nových forem elektronického publikování a komunikace informací, často se zdůrazňuje princip jejich otevřenosti („open access“, „open archives“), jež umožňuje jejich snadnou integraci či propojování na základě budování metadatových databází provozovateli služeb a také nové použití již uložených informací
- ***Institucionální repozitář*** – moderní termín, který se zrodil a začal rozšiřovat zhruba po roce 2000 v souvislosti s intenzivním zakládáním institucionálních repozitářů (IR); ve své podstatě jde ale o variantu (synonymum) základních výše definovaných termínů „elektronický archiv“, „digitální archiv“ nebo „archiv elektronických tisků“, která zahrnuje místo základního slova „archiv“ slovo „repozitář“ (ten je z pohledu obecného modelu komponentou elektronického archivu)
- ***Digitální [Elektronické] muzeum*** (angl. Digital [Electronic] Museum) - termín pro systém, který zajišťuje ukládání a zpřístupňování digitalizovaných nebo digitálních muzejních informací, tj. především exponátů muzeí v digitální obrazové, zvukové, multimediální aj. formě
- ***Digitální [Elektronická] galerie*** (angl. Digital [Electronic] Gallery) - termín pro systém, který zajišťuje ukládání a zpřístupňování digitálních galerijních informací, tj. především uměleckých exponátů galerií v digitální obrazové formě (galerie výtvarného umění spadají pod muzea); existují však také elektronické systémy pro obchodování různého zboží přes internet, které se mohou nazývat také elektronickými galeriemi (například pro prodej květin aj.)
- ***Informační brána*** (angl. Information Gateway) - obecný a velmi frekventovaný termín pro systém, zajišťující síťový přístup k různým typům a zdrojům informací určitého oborového nebo předmětového vymezení, jeho podstatou je vyhledávací systém a katalog popisující propojované informační zdroje (katalog má povahu znalostní databáze); vyskytují se také jeho synonyma: „předmětová brána“, „oborová brána“ aj.

Od výše uvedených termínů je třeba odlišovat termín:

- ***Automatizovaná knihovna*** (Automated Library)

Tento termín je určen pro označení konvenčních knihovnických institucí zabývajících se zpracováním, ukládáním a zpřístupňováním především tradičních dokumentů, které jsou automatizované, tj. mají automatizovaný jeden nebo více procesů tradiční činnosti (akvizice a katalogizace dokumentů, výpůjční systém a některé další); terminologická databáze TDKIV [KTD, 2003] i národní soubor autoritních dat AUT (NK ČR) zahrnují jen preferovaný termín „automatizované knihovní systémy“ ve smyslu softwaru pro

knihovnické procesy; v rámci konvenčních knihoven mohou být budovány dílčí digitální knihovny či fondy digitalizovaných dokumentů z fondů těchto knihoven (například starých tisků apod.)

V posledních letech se často objevuje také termín:

- **Hybridní knihovna** (Hybrid Library)

Tento termín vyjadřuje „nutné“ spojení konvenční knihovny a jejích služeb s novodobými systémy digitálních informací (primárních i sekundárních); zajišťuje informace jak z tradičních informačních zdrojů, tak ze zdrojů digitálních či digitalizovaných; rozvíjejí se aplikace (software), které budou uživatelům zpřístupňovat informace pocházející z tradičního i digitalizovaného informačního fondu jednotným způsobem a přes jednotné rozhraní prostřednictvím počítačové sítě. Zpřístupňování tradičního dokumentu znamená v současné době užití buď tradiční meziknihovní výpůjční služby, nebo jedné z moderních forem tzv. „elektronického zpřístupňování dokumentů“ (EDD, Electronic Document Delivery); pro srovnání, v národní databázi autoritních záznamů AUT (NK ČR) je termín „hybridní knihovny“ v podřazeném vztahu k termínu „digitální knihovny“

2.4 Problematika definice digitálních knihoven a jejich funkcí

Termín „digitální knihovna“ je termínem problémovým, protože skrytě zahrnuje komplex vztahů mezi **elektronickými informačními fondy a knihovnami jako institucemi**. K současnému datu bylo různými autory (především pak počítačovými vědci a knihovníky) publikováno **velké množství jeho definic**, které se velmi různí podle toho, zda se autor definice přiklání k jednomu nebo druhému základnímu aspektu celého sousloví. Termín je dokonce některými americkými autory označen za **oxymóron**, tj. slovní spojení zdánlivě si odporujících slov (v tomto případě slov „digitální“ a „knihovna“). Americký autor **D. Greenberg** uvedl v této souvislosti v roce 1998 slavné sdělení (je často citováno): „Jestliže je nějaká knihovna knihovnou, pak nemůže být digitální. Jestliže je nějaká knihovna digitální, pak to nemůže být knihovna“ [GREENBERG, 1998, s. 108]. Významný americký odborník v oblasti počítačové vědy se zaměřením na digitální knihovny **William Y. Arms** (prof. Cornellovy univerzity, USA) definuje digitální knihovnu „jako spravovaný fond informací spolu s odpovídajícími službami, přičemž informace jsou uloženy v digitální podobě a jsou dostupné prostřednictvím sítě ...“ (klíčovou částí definice je sdělení, že informace jsou spravované) [ARMS, 2000b, chapter 1]. Tvůrce slavné komplexní novozélandské digitální NZDL **Ian H. Witten** (prof. Waikatské univerzity, Nový Zéland) definuje digitální knihovnu „jako cílený fond digitálních objektů, zahrnující objekty textové, vizuální a zvukové, spolu s metodami pro jejich zpřístupnění a získávání, stejně jako pro výběr, organizaci a uchovávání [WITTEN, 2002]. **Donald Waters** (Mellonova nadace, USA) uvádí, že digitální knihovny jsou organizace (instituce) zahrnující zdroje (včetně specializovaného personálu), které umožňují provádět výběr, strukturování a zpřístupňování fondů digitálních materiálů, tyto materiály dále distribuovat, udržovat jejich integritu a dlouhodobě uchovávat -- to vše s ohledem na snadné a ekonomické využití určitou komunitou či skupinou komunit uživatelů [WATERS, 1998].

Jak poznamenávají mnozí odborníci, termín „digitální knihovna“ je nadále předmětem neustálých diskusí, definice je tedy stále ve vývoji a může se i nadále proměňovat.

Podle americké autorky **Christiny Borgmanové** (prof. Kalifornské univerzity, USA), která provedla rozsáhlou analýzu pojmu „digitální knihovna“ ve světovém kontextu, lze termín „digitální knihovna“ definovat v zásadě dvojím možným způsobem (jde o dva možné typy výkladu tohoto pojmu) [BORGMAN, 2003a, s. 42; BORGMAN, 2003b, s. 257]:

1. Digitální knihovny jsou souborem elektronických zdrojů a technických a technologických prostředků pro vytváření, vyhledávání a využívání informací (zahrnují data a metadata popisující různé aspekty dat); v tomto smyslu jsou digitální knihovny **rozšířením a obohacením systémů pro ukládání a vyhledávání informací (tedy databází)**, které manipulují s digitálními daty na jakémkoliv médiu (text, obraz, zvuk, video aj.) a existují v distribuovaných sítích
2. Digitální knihovny lze pojímat i jako **širší systémy sloužící určité komunitě uživatelů**, ve kterých probíhá interakce mezi jednotlivci či skupinami; využívá se různých dat, informací, znalostních zdrojů a systémů; v tomto smyslu jsou digitální knihovny **extenzí, obohacením a integrací různých informačních institucí jako fyzických míst (tedy knihoven, muzeí, archivů, škol a jiných korporací)**

Digitální knihovny by měly zajišťovat následující **základní funkce**:

- **zjišťování, výběr a získávání** digitálních objektů (informací); specifickým případem je proces elektronického podávání a metadatového popisu dokumentu do elektronického archivu, zajišťovaný samotnými autory či editory publikací (proces auto-archivace)
- **identifikace a zpracování** digitálních objektů (přidělování identifikátorů – v optimálním případě trvalých identifikátorů, tvorba nebo přebírání metadat, katalogizace, věcná indexace ruční povahy nebo zcela automatická indexace v případě úplných textů, konverze metadat apod.)
- **uložení**, resp. i **dlouhodobá archivace a ochrana** digitálních objektů v digitálních skladištích (repozitářích); **údržba a zajištění bezpečnosti dat**
- **vyhledávání, rozšiřování a zpřístupňování** digitálních informací koncovým i strojovým uživatelům DK (libovolným nebo oprávněným) včetně zajištění kontroly přístupu k objektům; může jít o služby známé již z dřívějších automatizovaných informačních systémů jako služby typu SDI (nazývané také služby typu Alerts, nebo moderní služby typu RSS (1.0, 2.0) nebo Atom), dodávání záznamů na základě personálních profilů uživatelů, statistiky, citační služby, bibliometrické analýzy nebo nové služby propojování informací uložených v digitální knihovně s informacemi z jiných systémů; specifickým případem služeb DK je volné poskytování metadat jiným informačním službám prostřednictvím komunikačních protokolů (zejména pak protokolu OAI-PMH)

Vedle základních funkcí může být digitální knihovna rozšířena o **funkce další** podle toho, v jakém prostředí působí. Například v oblasti univerzit jsou budovány digitální archivy disertací, které také zajišťují důležitou funkci **automatického podávání (předkládání) prací** v elektronické formě (popřípadě i v návaznosti na jiné školní systémy). Některé z těchto archivů mají i další funkci, totiž funkci **prodeje úplných textů** (může jít o výběr). Prodej informací nebo přístupu k informacím je jinak typický pro digitální knihovny komerčních vydavatelů.

V rámci budování či zdokonalování současných digitálních knihoven se řeší a rozhoduje řada dále **uvedených problémových okruhů**:

- problematika **obsahu (předmětu)** zpracování, uložení a/nebo dlouhodobé archivace digitálních objektů (informačních objektů) v digitálních knihovnách, archivech apod.; pojetí informačního objektu v modelu OAIS je osvětleno v části 2.7; problematika typologie informačních zdrojů je podrobněji pojednána v další podkapitole; otázky jednoznačné identifikace zdrojů a trvalých identifikátorů a problematika metadat jsou výjimečně rozvedeny až v 5. kapitole této práce

- problematika **technického zabezpečení** se týká zejména zabezpečení kvalitní hardwarové základny provozu digitální knihovny; patří k ní (obecně) jednak **WWW server**, určený ke komunikaci a publikování informací uložených v repozitáři (databázi) digitální knihovny či archivu, jednak **databázový server** k uložení digitálního fondu; servery mohou běžet pod různými operačními systémy (UNIX, Linux, Windows aj.) a měly by mít dostatečnou paměťovou kapacitu, zejména pokud jde o digitální fond
- problematika **technologie zpracování, ukládání a zpřístupňování digitálních objektů**, která se týká koncepce, architektury digitální knihovny, uložení a organizace digitálních objektů včetně metadat v celkové struktuře digitální knihovny, protokolů pro komunikaci objektů uvnitř i vně systémů, bezpečnosti dat apod. s návazností na technické zabezpečení; trendem současnosti, pokud jde o základní **softwarové vybavení**, je využívání tzv. **otevřených zdrojů** (Open Sources, tj. programů s přístupným zdrojovým kódem), zejména pak těch, které jsou k dispozici zdarma (Free Software); patří k nim databázový program MySQL, PostgreSQL, miniSQL aj.; k tvorbě vlastních aplikací, využívajících rozhraní WWW, je užíváno řady prostředků (programovací jazyk Perl, standard CGI=Common Gateway Interface, populární PHP, nebo favorizované nástroje JSP, Java Servlets a J2EE); k aplikaci skriptů je potřebná instalace kompatibilního WWW serveru. Doporučován je zejména volně dostupný Apache Web Server, v rámci projektu Jakarta Apache jsou k dispozici volně dostupné nástroje, jako program Tomcat a jeho aplikace Cocoon (k publikování informací na WWW); charakteristika nejpoužívanějších programů v oblasti archivů elektronických tisků je obsažena v konferenčním příspěvku [KREJČÍŘ, 2006]; elektronické archivy ve vědě a zejména pak rozsáhlé digitální knihovny národního charakteru se zaměřením na dlouhodobou archivaci budují v současnosti své systémy v souladu s modelem OAIS, tj. v souladu s mezinárodním standardem [ISO 14721:2003]
- **organizační problémy** provozu digitálních knihoven včetně otázek funkcí a otázek finančního zabezpečení
- složitá problematika **autorsko-právní** ochrany digitálních děl (zahrnuje otázky autorských práv k dílu v digitální podobě, vztahu mezi autory a provozovateli digitálních knihoven a eventuálně i vydavateli komerčními a vztah mezi autory a uživateli jejich děl apod.)
- problematika **mezinárodní kooperace** budování digitálních knihoven (společné budování digitálních knihoven, archivů apod. v celosvětovém rámci, jejich postupné a účelné propojování – integrace v síťovém prostoru; budování národních a mezinárodních metadatových registrů aj.)

2.5 **Problematika typologie digitálních objektů (elektronických zdrojů)**

Předmětem prvních digitálních knihoven, resp. sbírek (kolekcí) digitálních informací, byly **textové informační zdroje**. Šlo především o textové soubory digitalizovaných technických zpráv a také knižních materiálů, zejména starých tisků, které se původně vyskytovaly v papírové podobě. S vývojem informačních technologií přibývaly další typy objektů, a tak v současné době může „digitální knihovna“ zpřístupňovat jakýkoli typ informace (informačního zdroje či informačního objektu).

Třídění nebo typologie elektronických zdrojů je složitou otázkou, jak dokladují dlouhodobé publikované debaty odborníků v posledních 13-15 letech. Autorka této práce se touto problematikou podrobně zabývala v publikovaném článku z roku 1988, který byl v roce 2005 aktualizován [BRATKOVÁ, E. 2005a]. Zahrnuje též komentář ke komplikované situaci zařazování elektronických zdrojů do třídění (či typologie) všech zdrojů a třídění těchto zdrojů

samotných v oblasti knihovnictví (ve formátech typu MARC), v kontextu této práce, věnované digitálním knihovnám a metadatům, bude však užitečné krátce připomenout dvě typologie elektronických zdrojů, které se využívají v metadatových popisech, a to typologii zpráv MIME (včetně související typologie datových formátů) a typologii informačních zdrojů Dublin Core [BRATKOVÁ, E. 2005a, kap. 2 a 4].

2.5.1 Typologie elektronických zpráv v rámci internetové normy MIME

První typologie zahrnuje „zprávy“ přenášené v rámci protokolů TCP/IP. Je ustanovena internetovým standardem označovaným zkratkou **MIME** (Multipurpose Internet Mail Extensions, tj. víceúčelová rozšíření internetové pošty). Celý standard představuje soubor textových dokumentů RFC 2045 - RFC 2049 přístupných v síti Internet. MIME je standardem, který rozšiřuje původní standardy RFC 821 - RFC 822, jež byly věnovány přenosu jednoduchých textových zpráv přes elektronickou poštu. MIME zajišťuje přepravu zpráv složitějšího charakteru - textů s diakritikou, obrázků, zvuků apod. prostřednictvím stávajícího poštovního systému (podstatnou roli ovšem hraje v tomto směru způsob kódování složitějších zpráv do základního ASCII kódu). Typologie zpráv je definována přesně ve 2. části standardu - RFC 2046 [RFC 2046, 1996]. V jeho názvu se užívá termínu „**typy médií**“ (media types) ve smyslu typů zpráv jako zprostředkovatelů přenosu informačního obsahu. Typologie je zejména pro provoz moderních služeb internetu (WWW aj.) velmi užitečná a uplatňuje se ve větší či menší míře v jednotlivých informačních systémech internetových, ale i knihovnických zabývajících se zpracováním zdrojů, a to jak pokud jde o formát, tak pokud jde o typ či charakter přenášené informace.

Hodnoty typů zpráv jsou obsaženy v části **záhlaví zprávy** (v hlavičce), tj. řádce označené návěšním „Content-Type“ (typ dat v obsahu zprávy). Specifikuje přesně charakter obsahu zprávy pomocí definovaného typu a podtypu s případnou doplňkovou informací (o jménu souboru, znakové sadě apod.). Podtypy pak prezentují konkrétní formáty zpráv a budou hrát důležitou roli při vyhledávání informací. V rámci předepsané syntaxe vypadá zápis informace v hlavičce zprávy následujícím způsobem (příklady):

Content-type: text/html; charset=ISO-8859-2

Content-type: image/gif

Content-type: application/msword; name="clanek.doc"

Content-type: model/vrml

RFC 2046 definuje dnes celkem 6 jednoduchých a 2 složené typy zpráv, dat či informací. V jejich rámci je k dnešnímu dni definována velká řada dílčích podtypů prezentujících známé i méně známé formáty přenášených zpráv. Typy a podtypy (stejně tak znakové sady, typy přístupů aj.) jsou schvalovány úřadem IANA (The Internet Assigned Numbers Authority) a registrovány na jeho veřejném WWW serveru [Internet Assigned, 2007].

K jednoduchým typům zpráv patří [RFC 2046, 1996]:

1. **text** (text) - typ k posílání textových (grafických písemných) informací; základním podtypem je formát „plain“ (čistý neformátovaný text), známým podtypem je dnes frekventovaný formát „html“ (text obohacený o značky jazyka HTML); textové informace vyžadující k percepci speciální program jsou zařazovány k typu „application“
2. **image** (obraz) - typ k posílání obrazových (ikonických nebo ideografických) informací (včetně animovaných), k jejichž zobrazení je třeba specifického zařízení (grafická obrazovka, grafická tiskárna aj.); k nejznámějším podtypům patří formát „jpeg“ a „gif“

3. **audio** (zvuk) - typ k posílání zvukových (auditivních) informací, k jejichž výstupu je potřeba příslušného přehrávacího zařízení; základním definovaným podtypem je formát „basic“
4. **video** (video) - typ k posílání pohyblivých obrazových informací, k jejichž zobrazení je třeba rovněž specifického přehrávacího zařízení; základním podtypem je formát „mpeg“
5. **application** (aplikace) - specifický typ k posílání jiných typů informací, zpravidla buď binárních dat nebo informací, které je nutné zpracovat pomocí nějaké aplikace (programu), aby byly čitelné pro uživatele; definovány jsou dva základní a velké množství specializovaných podtypů formátů; k základním patří: 1. formát „**octet-stream**“ (obsahem zprávy jsou binární data) a 2. formát „**postscript**“ (obsahem zprávy je postscriptový dokument, který je čitelný pouze pomocí speciálního programu); **dalších specializovaných podtypů** je v současné době velké množství - jejich přehled je k dispozici v internetovém dokumentu „Typy médií MIME“ [Internet Assigned, 2007]; v jejich rámci je zařazena i řada podtypů prezentujících v jádru textové informace (některé formáty se i opakují), podstatné pro jejich zařazení mezi typ „aplikace“ je ovšem jejich zpracování pomocí nějakého programu (viz formát „msword“, „pdf“, „sgml“ nebo i formát „marc“ strukturovaný podle normy ISO 2709)
6. **model** (model) - typ k posílání troj- a vícerozměrných systémů, ve kterých lze zavést pravoúhlou soustavu souřadnic; model se skládá z jednoho nebo více objektů, které se pak skládají z prvků, jež mají mezi sebou definovány vztahy; místo slova model se užívá i hovorového termínu „virtuální realita“; tento typ zpráv byl v normě RFC 2046 uveden původně jako experimentální, nyní je jeho popis stanoven normou RFC 2077 [RFC 2077, 1997]; k základním podtypům patří formát „vrml“

Novým typem, který přibyl do seznamu IANA, je typ „example“. Je určen pro potřeby dokumentace. Nemá definované žádné formáty, a proto nebude blíže komentován.

K složeným (kompozitním) typům zpráv, které obsahují více dílčích zpráv, patří:

1. **multipart** (multipart) - typ, který obsahuje několik dílčích zpráv; k základním podtypům patří: 1. formát „multipart/mixed“ (obsahuje dílčí nezávislé zprávy ve stanoveném pořadí), 2. formát „multipart/alternative“ (obsahuje dílčí zprávy se shodnými informacemi, ale v různém tvaru), 3. formát „multipart/digest“ (obsahuje zprávu nebo více zpráv tvořených posloupností jiných dílčích zpráv), 4. formát „multipart/parallel“ (obsahuje dílčí zprávy bez ohledu na jejich řazení) a 5. formát „multipart/signed“ a „multipart/encrypted“ (tzv. bezpečné zprávy složené ze základní dílčí zprávy a elektronického podpisu nebo ze šifrované zprávy)
2. **message** (zpráva) - typ, který umožňuje poslat: 1. zprávu jako tělo jiné zprávy (vnořená zpráva) - podtyp „message/rfc822“, 2. poslat dlouhou zprávu jako několik kratších - podtyp „message/partial“ a 3. poslat informaci o zprávě uložené na nějakém serveru - typ „message/external“

Většina definovaných typů MIME koresponduje s typologií sociálních, příp. fyzikálních informací z hlediska způsobu jejich prezentace a percepce. Nicméně, typologie má význam především z hlediska způsobu uspořádání dat při jejich přenosu v počítačové síti, a proto se také často uplatňuje v dále zmiňovaných systémech při dělení informačních zdrojů podle formátu. Základní rozdělování informačních zdrojů se ubírá mnohdy jinými směry.

2.5.2 Typologie informačních zdrojů standardu Dublin Core

Jednou z nejvýznamnějších metadatových specifikací dneška je specifikace označovaná zkratkou **DC** (Dublin Core), v úplném tvaru pak **Dublin Core Metadata Element Set** (Soubor metadatových prvků Dublin Core). U jeho zrodu v roce 1995 stáli pracovníci výzkumného oddělení OCLC, kteří ve spolupráci s dalšími odborníky z celého světa založili mezinárodní virtuální společenství **DCMI** (Dublin Core Metadata Initiative) [Dublin Core, 2005]. Specifikace DC se během let stala mezinárodní normou. Je využívána v řadě systémů pro registraci elektronických zdrojů po celém světě. V této části práce nás bude zajímat jen jediný moment celého formátu, a to jeho typologie zdrojů nahlížená stručně v historickém kontextu.

Typologie informačních zdrojů byla připravována odborníky více jak 10 let. Je určena pro jeden z 15 základních prvků specifikace DC – prvek **Resource Type** (Typ zdroje). Jiný prvek **Format** (Formát) je určen pro typologii datových formátů.

Základní typologii byla věnována velká pozornost od samého vzniku specifikace DC. Podílela se na ní komunita odborníků z oblasti knihovnictví i počítačové vědy, jednotlivé návrhy a problémy byly diskutovány na pracovních seminářích a v elektronické konferenci. Podívejme se dále na její jednotlivé verze, od první úplné z července 1997 [Dublin Core Resource Types, 1997] až po zatím poslední z roku 2004 [Dublin Core, 2004].

Třístupňová typologie (tehdy jen elektronických) informačních zdrojů DC (verze z roku 1997) zahrnovala následující typy:

<u>Text</u>	<u>Image</u>
Text.Abstract	Image.Moving
Text.Advertisement	Image.Moving.Animation
Text.Article	Image.Moving.Film
Text.Correspondence	Image.Photograph
Text.Correspondence.Discussion	Image.Graphic
Text.Correspondence.Email	
Text.Correspondence.Letter	<u>Sound</u>
Text.Correspondence.Postcard	Sound.Ambient
Text.Dictionary	Sound.Effect
Text.Form	Sound.Music
Text.Homepage	Sound.Narration
Text.Homepage.Organizational	Sound.Speech
Text.Homepage.Personal	
Text.Index	<u>Software</u>
Text.Manual	Software.Executable
Text.Manuscript	Software.Source
Text.Minutes	
Text.Monograph	<u>Data</u>
Text.Pamphlet	Data.Numeric
Text.Poem	Data.Spatial
Text.Proceedings	Data.Spectral
Text.Promotion	Data.Statistical
Text.Serial	Data.Structured-Text
Text.Serial.Journal	
Text.Serial.Magazine	<u>Interactive</u>
Text.Serial.Newsletter	Interactive.Chat
Text.Serial.Newspaper	Interactive.Games
Text.TechReport	Interactive.Multimedia
Text.Thesis	Interactive.VR
Text.Thesis.Doctoral	
Text.Thesis.Masters	

Verze z roku **1997** byla až třístupňová. První úroveň představovala 6 základních typů informačních zdrojů, jež vycházely z typu informace v nich obsažené. Některé kategorie korespondovaly se známou typologií sociálních informací (**text**, **obraz**, **zvuk**), popř. také s typologií MIME. Reprezentantem hybridního typu informace byl typ zdroje **program** (software), typ **data** a zejména novodobý reprezentant síťových zdrojů **interaktivní zdroj** (interactive). Druhá a třetí úroveň typologie zahrnovala podtypy informačních zdrojů, vymezených souhrnem znaků, k nimž náležely minimálně jejich funkce a cílové určení a forma uspořádání informací. Nejrozsáhlejší dělení bylo definováno pro typ **text**, v jehož rámci se vyskytovaly jak podtypy známé z tradiční typologie textových dokumentů (monografie, seriál, technická zpráva, korespondence aj.), tak podtypy novodobé (formulář, domovská stránka aj.). Výčet podtypů nebyl úplný, pro nové verze se předpokládala další úprava a doplňování. Totéž platilo i pro podtypy dalších speciálních typů zdrojů. Komplikovaná situace byla ve vymezení typů **data** (soubory dat) a **interaktivní zdroje**, které byly doplňovány i v souvislosti s rozvojem informačních technologií.

Další návrh typologie informačních zdrojů DC (tentokrát se ale už šlo nad rámec elektronických zdrojů) pocházel z roku **1998** (připraveny byly v tomto roce celkem 4 verze, poslední úprava byla schválena 23. 10. 1998). Roční práce řady odborníků seskupených kolem speciální pracovní skupiny „Resource Type & Format Working Group“ doznala značných změn. Šlo o následujících 8 typů informačních zdrojů [Dublin Core, 1998]:

1. **text** (text) - zdroj, jehož obsah je určen především pro čtení (např. knihy, dopisy, básně, noviny aj.); k tomuto typu zařazovali tvůrci typologie i faksimile nebo obrázky textů
2. **obraz** (image) - zdroj, jehož obsah představuje symbolickou vizuální (zakovou), resp. obrazovou reprezentaci, ale jinou než u textu (např. obrázky, fotografie fyzických objektů, obrazy, kresby, animace, filmy, videozáznamy, diagramy, mapy, hudební notace)
3. **zvuk** (sound) - zdroj, jehož obsah je určen především k poslechu (např. hudba, projev, záznam zvuku)
4. **soubor dat** (dataset) - zdroj, jímž se rozumí strukturované informace kódované v seznamech, tabulkách, databázích apod., které se normálně vyskytují ve formátu pro přímé strojové zpracování (např. tabulkové procesory, databáze, data geografických systémů aj.); nestrukturované číselné nebo slovní údaje jsou považovány za typ „text“
5. **program** (software) - počítačový program ve zdrojové nebo kompilované formě, který je dostupný trvale pro instalaci na jiných počítačích; programy, které vytvářejí interaktivní prostředí při komunikaci, je nutné zařadit k typu „interaktivní zdroj“
6. **interaktivní [zdroj]** (interactive) - zdroj vyžadující interakci uživatele, aby informace byly pochopeny, provedeny nebo využity (např. formuláře na webových stránkách, aplety, multimediální výukové předměty, konverzační služby, virtuální realita)
7. **událost** (event) - zdroj s časově omezeným výskytem (např. výstava, konference, představení aj.); metainformace o události (akci) nemusí identifikovat vyhledatelný zdroj, pokud její čas již vypršel a více se nevyskytuje
8. **fyzický objekt** (physical object) - trojrozměrné objekty nebo látky, které nelze považovat za texty, obrázky nebo jiné typy výše uvedené (např. osoba, počítač, velká pyramida, socha); digitální reprezentace nebo zástupce uvedených objektů musí tedy figurovat jako typ „obraz“, „text“ apod.

Novinkou této verze typologie (1998) bylo zařazení typu **událost** a **fyzický objekt**. V diskusi roku 1998 byly posuzovány i pojmy **složený (kompozitní) zdroj**, popř. **smíšený zdroj** a pojem **sbírka** (collection). Pro základní jednoduchou variantu „DC.Typy“ byly však tehdy

odmítnuty, a to především z důvodů jejich vyhledávání. Složené zdroje bylo výhodnější zařazovat k více dílčím typům (např. multimediálnímu programu s jedním URL bylo možné přiřadit 4 typy - zvuk, text, obraz a interaktivní zdroj). Sbírky (kolekce) zdrojů měly být přiřazeny spíše k jednomu základnímu dílčímu typu. Na místě bylo také uplatňování vazeb (relací), zvažovalo se i jejich případné uplatnění jako podtypů (např. Text.Sbírka).

Nejrozsáhlejší verze typologie DC všech informačních zdrojů byla přijata v srpnu 1999 [Dublin Core, 1999]. Zahrnuty byly (kurzivou jsou označeny tehdy nově navržené typy):

1. *Sbírka (Collection)*
2. Soubor dat (Dataset)
3. Událost (Event)
4. Obraz (Image)
5. Interaktivní zdroj (Interactive resource)
6. *Model (Model) jako abstrakce*
7. *Strana (Party) jako osoba nebo korporace*
8. Fyzický objekt (Physical object)
9. *Místo (Place) jako geografická entita*
10. *Služba (Service)*
11. Program (Software)
12. Zvuk (Sound)
13. Text (Text)

Novými položkami byl typ **Sbírka**, který byl specifikován jako agregát (soubor) dokumentů, dále typ **Model**, považovaný za symbolickou reprezentaci (abstrakci reálné věci), typ **Strana**, jíž mohla být fyzická osoba, organizace nebo instituce, a typ **Místo** ve smyslu geografické oblasti. Poslední tři jmenované typy, které více specifikovaly původně navržený typ „Fyzický objekt“, nebyly přijaty jednoznačně řadou odborníků, a proto byly z dalších návrhů (v letech 2000-2004) vyřazeny. Posledním novým typem, zařazeným do sestavy v roce 1999, byl typ **Služba**, jíž je míněn systém zajišťující jednu nebo více funkcí koncovým uživatelům (například MVS, autentizační služba, bankovní služba aj.).

Typologie informačních zdrojů DC z července 2000 znamenala návrat k menšímu počtu typů [Dublin Core, 2000]. Zařazeny byly:

1. Sbírka (Collection)
2. Soubor dat (Dataset)
3. Událost (Event)
4. Obraz (Image)
5. Interaktivní zdroj (Interactive resource)
6. Služba (Service)
7. Program (Software)
8. Zvuk (Sound)
9. Text (Text)

Své místo si uhájily typy **Sbírka** a **Služba**, po dlouhých debatách však byl (ale jenom na určitý čas) vyňat problémový a velmi diskutovaný typ **Fyzický objekt**. Do typologie zdrojů

byl tento typ opět zařazen až v červenci 2002, takže jejich celkový počet dosáhl počtu deseti typů zdrojů. Tento stav v podstatě potvrdila i další verze z února 2003 [Dublin Core, 2003a].

Nový směr v rozvoji typologie všech informačních zdrojů (elektronických i tradičních) znamenala až verze z 19. listopadu 2003 [Dublin Core, 2003b]. Její novinkou bylo (kromě formální úpravy celého seznamu) zařazení dvou nových typů, které znamenaly sémantické zjemnění typu **Obraz**. Šlo o typy:

1. **Stálý obraz (Still Image)**, který je definován jako statická vizuální reprezentace (malba, plán, mapa, kresba, grafický design aj.) a
2. **Pohyblivý obraz (Moving Image)**, který je definován jako série vizuálních reprezentací, které, když jsou ukazovány, vyvolávají dojem pohybu (animace, film, televizní program, video aj.)

Zařazením těchto dvou nových typů se typologie DC opět, pokud jde o obrazové informace, přiblížila typologii standardu MIME (část 2). Celkový počet položek této verze typologie všech informačních zdrojů DC zahrnoval tedy dvanáct typů.

Zatím poslední verze typologie informačních zdrojů DC pochází z června 2004 [Dublin Core, 2004]. Tato verze potvrdila 12 typů informačních zdrojů, změnou bylo pouze jejich abecední uspořádání (dle angličtiny) v seznamu:

1. Sběrka (Collection) - agregát jednotek, popisovaných jako skupina (popis dílčích jednotek je možný)
2. Soubor dat (Dataset) - informace kódované v definované struktuře, určené pro přímé strojové zpracování
3. Událost (Event) - zdroj s netrvalým a časově omezeným výskytem
4. Obraz (Image) - symbolická vizuální reprezentace jiná než text (obraz v elektronické či fyzické formě)
5. Interaktivní zdroj (Interactive resource) - zdroj, který vyžaduje interakci s uživatelem
6. Pohyblivý obraz (Moving Image) - série vizuálních reprezentací, které, když jsou ukazovány, vyvolávají dojem pohybu
7. Fyzický objekt (Physical Object) - neanimovaný trojrozměrný objekt (předmět nebo látka)
8. Služba (Service) - systém zajišťující jednu či více funkcí pro koncové uživatele
9. Program (Software) - počítačový program ve zdrojové či kompilované formě
10. Zvuk (Sound) - zdroj, jehož obsah je určen především k poslechu
11. Stálý obraz (Still Image) - statická vizuální reprezentace
12. Text (Text) - zdroj, jehož obsahem jsou především slova ke čtení

V systému formátu DC se uplatňuje i další typologie síťových elektronických zdrojů, která má důležitý doplňkový význam pro jejich vyhledávání. Formálním hlediskem se v tomto případě stává konkrétní formát zdroje (uspořádání dat) a je jen logické a rozumné, že tvůrci DC sáhli v tomto případě po známém internetovém standardu typologie elektronických zpráv MIME (viz charakteristika v předchozí části) v jeho poslední verzi RFC 2045 - RFC 2049. Celkový seznam schválených formátů, který udržuje a neustále doplňuje úřad IANA (Internet Assigned Numbers Authority) [Internet Assigned, 2007], je k dnešnímu dni již poměrně dlouhý, v rámci již existujících provozů se jich zatím však užívá méně.

V následujícím seznamu jsou uvedeny vybrané (ukázkové) typy zpráv a jejich formátů stanovené v rámci MIME (obě hodnoty odděluje syntaktické lomítko):

text/plain
text/richtext
text/html
text/sgml
text/xml
image/jpeg
image/gif
image/tiff
audio/basic
video/mpeg

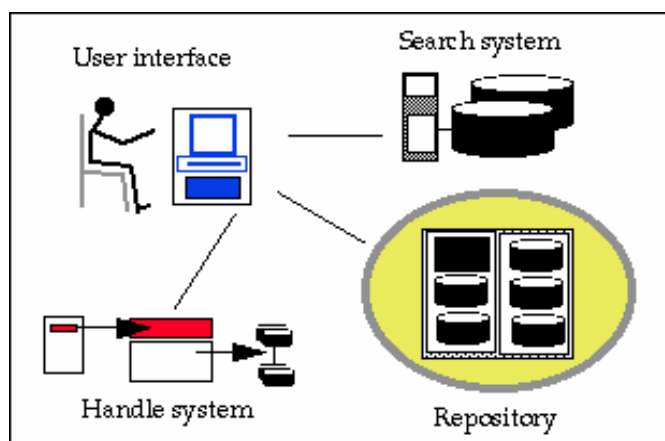
application/postscript
application/rtf
application/wordperfect5.1
application/pdf
application/powepoint
application/msword
application/pgp-encrypted
application/pgp-signature
application/marc
model/vrml

2.6 Architektura digitální knihovny podle amerického modelu „Kahn-Wilensky“

Digitální knihovny nebo elektronické archivy, provozované od 90. let 20. stol. na území USA, byly v řadě případů budovány na základě výchozí obecné architektury (modelu či rámce), jejíž základy byly připraveny v rámci výzkumných programů a projektů v letech 1993-1995. Podle konkrétních tvůrců tohoto modelu Roberta Kahna a Roberta Wilenského se tato architektura někdy označuje jako „**architektura Kahn-Wilensky**“ [KAHN, 1995; KAHN, 2006]. Významný byl v tomto kontextu především původní projekt digitální knihovny pro digitalizované technické zprávy z oboru počítačové vědy z fondů různých amerických univerzit (**Computer Science Technical Reports Project**, CSTR nebo CS-TR), <http://www.cnri.reston.va.us/cstr.html>, který vedla Korporace pro národní výzkumné iniciativy se sídlem v Restonu (Corporation for National Research Initiatives, CNRI, <http://www.cnri.reston.va.us/>) ve spolupráci s dalšími americkými institucemi. CNRI úzce spolupracovala také se skupinou IETF (Internet Engineering Task Force). Projekt byl financovaný z prostředků Agentury pro rozvojové výzkumné projekty řízené Ministerstvem obrany USA (Defence Advanced Research Projects Agency, DARPA) [KAHN, 2006, s. 115].

Komponentami architektury digitální knihovny modelu „Kahn-Wilensky“ jsou (viz obr. 2/1):

- ❑ **Uživatelské rozhraní** (User Interface), v jehož rámci koncový uživatel, knihovník nebo správce využívá systém digitální knihovny prostřednictvím prohlížeče a návazných klientských služeb



Obr. 2/1: schéma základních komponent digitální knihovny v modelu „Kahn-Wilensky“ [ARMS, 1997]

- ❑ **Repozitář** (Repository) pro ukládání **digitálních objektů** (zahrnují data i metadata). Repozitáře (specifická digitální úložiště či digitální skladiště pro digitální objekty) mohou být moderního typu, zajištěného příslušnou svébytnou softwarovou aplikací (v současnosti je vznášen požadavek na budování spolehlivých či důvěryhodných úložišť), může jít ale také o databáze naskenovaných (digitalizovaných) dokumentů, nebo jimi mohou být webové servery. K přenosu informací z repozitáře je využíván (obecně) repozitářový přístupový protokol (RAP, angl. Repository Access Protocol)
- ❑ **Identifikační systém** (angl. Handle System) pro generování či přidělování, správu a zejména směrování (angl. resolution) jednoznačných identifikátorů digitálních objektů (HDL, URN aj.)
- ❑ **Vyhledávací systém** (Search System) pro vyhledávání informací v digitální knihovně; vyhledávací systém zahrnuje indexy, reprezentující buď jen popisná metadata (informace o digitálních objektech), a/nebo také z úplné texty, jsou-li předmětem zpracování

Řadu principů obecného rámce (architektury) projektu CSTR je možné částečně implementovat v rámci WWW (korespondují se zásadami vyhlášenými internetovou organizací IETF), například:

- digitální objekty (viz dále) včetně meta-objektů mohou být uloženy ve webových (elektronických) archivech
- identifikační server (Handle Server) podporuje webová URL (je schopen je směrovat)
- termín URC (Uniform Resource Characteristic) pocházející z IETF je formou meta-objektu
- identifikátor (Handle) je specifickou formou URN (Uniform Resource Name) - platnou v USA (jejich struktura a syntax se však liší)

2.6.1 Identifikátory a systémy pro jejich správu a směrování

Pro komunikaci (ukládání, zpracování, uložení, vyhledávání, směrování) digitálních objektů v digitálních knihovnách hrají důležitou roli jejich **identifikátory (angl./amer. Handles)**. Jde jinak o jména, která by:

- měla být nezávislá na lokaci digitálního objektu
- měla být jedinečná v globálním měřítku
- měla být trvalá z hlediska času (angl. persistent)
- byla generována automaticky (automatické přidělování těchto jmen)

Systém pro správu identifikátorů (Handle Management System) by měl zajistit následující úlohy:

- rychlé směrování identifikátorů (angl. resolution) v rámci vyhledávacích procesů
- tvorbu kopií (replik) digitálních objektů a jejich ukládání ve vyrovnávací paměti (caching)
- decentralizovanou správu a řízení celého systému
- kontrolu změn digitálních objektů

Současné aktivity jsou zaměřeny na podporu identifikátorů ze strany standardních uživatelských rozhraní.

Identifikátory (Handles) jako unikátní jména měly v modelu „Kahn-Wilensky“ následující strukturu. Příklady (jde o původní struktury z 90. let 20. stol.) [ARMS, 1997]:

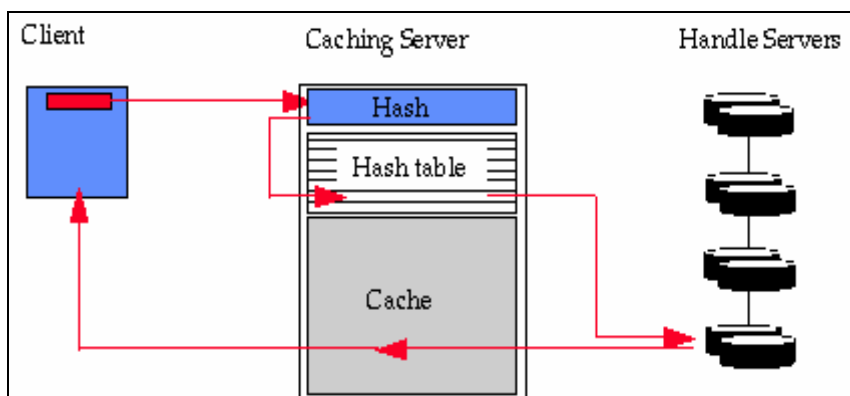
loc.ndlp.amrlp/3a16116

cnri-1/1995.02.12.16.42.21;9

Údaj před lomítkem (prefix) představuje administrativní jednotku, která přiděluje jména, údaj za lomítkem (sufix) je lokálním jedinečným identifikačním číslem digitálního objektu. Administrativní jednotky mohou být zřizovány v hierarchické struktuře.

Systém pro správu identifikátorů by měl být veřejně dostupný a měl by mít následující části:

1. **Generátory identifikátorů (Handle Generators)**, které na základě stanoveného algoritmu (programu) vytvářejí globálně jednoznačné identifikátory, jež mohou být spojeny s digitálními objekty
2. **Identifikační (směrovací) servery (Handle Servers)**, které slouží 1. k ukládání identifikátorů včetně příslušných lokačních či jiných informací (údajů) do databáze směrovacího serveru (směrovací služby), 2. ke zpracování dotazů klienta týkajících se směrování identifikátorů a 3. k vrácení údajů o lokaci směrem ke klientovi
3. **Adresář identifikačních serverů (Handle Server Directory)**, který zajišťuje správu jejich seznamu



Obr. 2/2: schéma identifikačního systému digitální knihovny [ARMS, 1995b]

Schéma na obr. 2/2 ukazuje zcela obecný způsob využití identifikačních serverů ke směrování identifikátorů a poskytování odkazů (referencí) na digitální objekty, které identifikátory identifikují.

Hašovací tabulka (Hash table) je využívána ke zjištění a označení identifikačního serveru (Handle server), který vlastní informace spojené s určitým identifikátorem. Existují 3 běžné konfigurace:

- klient se může spojit se serverem pro vyrovnávací paměť (Caching Server), který zná hašovací algoritmus (Hash), má kopii hašovací tabulky a má ve vyrovnávací paměti (Cache) zaznamenány naposledy využívané identifikátory (jde o konfiguraci zobrazenou na 2/2)
- klientský program může znát hašovací algoritmus a vlastnit kopii hašovací tabulky. Může udržovat vlastní vyrovnávací paměť. Jde o ekvivalent klienta fungujícího podobně jako jeho vlastní server pro vyrovnávací paměť

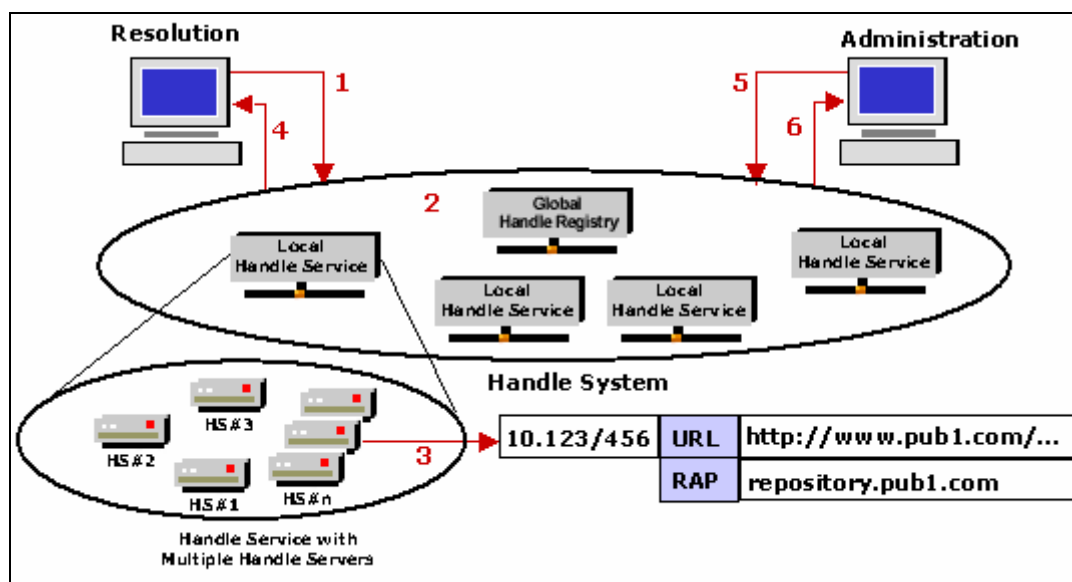
- klient se může spojit přímo s jakýmkoliv identifikačním serverem. Server provede vyhledání hašovací tabulky a předá dotaz dále příslušnému počítačovému serveru

Identifikační servery vracejí buď všechny uložené údaje nebo jenom dílčí soubory údajů.

Identifikátory a systémy pro jejich správu byly navrženy pro využívání v digitálních knihovnách, ale lze je využívat i pro ukládání větší skupiny údajů, například URL, e-mailové adresy apod.

Řada digitálních archivů (například v aplikaci DSpace.), zejména z USA, zavedla přidělování trvalých identifikátorů typu **Handle (HDL)**, jehož model byl připraven v rámci systému „**The Handle System**“ (<http://www.handle.net>). Byl založen a je rozvíjen a spravován organizací CNRI (Corporation for National Research Initiatives, <http://www.cnri.reston.va.us/>). Struktura tohoto trvalého identifikátoru (**HDL**) zahrnuje v základu dvě hlavní části, které jsou odděleny syntaktickým lomítkem (/). Část před lomítkem identifikuje jednoznačně instituci přidávající identifikátory, část za lomítkem identifikuje konkrétní digitální objekt. Globální server CNRI spravuje databázi korporací (Naming Authorities), které přidělují identifikátory v rámci své působnosti. V rámci lokálního serveru jsou pak digitálním objektům přidělována konkrétní HDL (americká „URN“) a zároveň je zajišťována a neustále aktualizována vazba mezi nimi a příslušnými URL. Program k zajištění uvedených procesů lze stáhnout ze sídla systému „The Handle System“ (Local Handle System Server-JAVA). Součástí celého programového balíku je i plug-in (Handle System Resolver Plug-In), který zajišťuje směrovací službu (k cílovému vyhledání samotného objektu).

Americký systém „The Handle System“ se stal také základem mezinárodního identifikačního systému DOI (<http://www.doi.org>), který je hojně využíván komerčními vydavateli (proxy server pro směrování identifikátorů DOI je k dispozici veřejně na URL: <http://dx.doi.org/>).



Obr. 2/3: schéma architektury globálního identifikačního systému a jeho operací (verze z 90. let 20. stol.) [převzato z původního WWW CNRI]

Identifikační systém by měl potřebné funkce (tvorba identifikátorů i jejich směrování) zajišťovat **distribuovaným způsobem**. Obr 2/3 znázorňuje (ve starší verzi z konce 90. let 20. stol.) model **globálního identifikačního systému**. Tvorbu identifikátorů i jejich směrování zajišťují vyhrazení klienti (webové prohlížeče apod.). Ve všech případech je

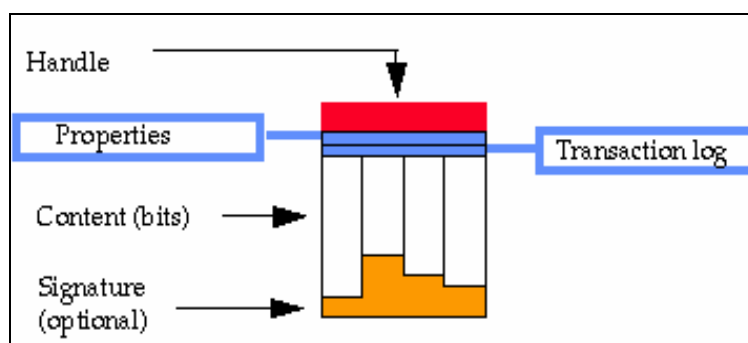
komunikace s identifikačním systémem zabezpečována pomocí protokolů tohoto identifikačního systému (existují i specifické implementace v rámci protokolů pro komunikaci ze strany klienta zajišťujícího tvorbu identifikátorů (Administration) i klienta zajišťujícího směrování identifikátorů (Resolution)). V systému dochází k následujícím operacím:

1. V rámci **směrování identifikátorů (Resolution)** potkává klient (proces je na obr. 2/3 označený číslem 1) identifikátor (například *10.123/456*) na Internetu, popř. Intranetu v podobě hyperodkazu nebo jiného typu odkazu. Klient odesílá identifikátor ke směrování k identifikačnímu (směrovacímu) systému (může k tomu dojít přímo na straně klienta nebo prostřednictvím proxy serveru).
2. **Vlastní identifikační (směrovací) systém** (na obr. 2/3 pod číslem 2) je tvořen skupinou identifikačních (směrovacích) služeb. Každá služba může být fyzicky i logicky distribuována na libovolném počtu serverů a každý server může být replikován (kopie). Jedna ze služeb - Globální registrace identifikátorů (Global Handle Registry) - zajišťuje správu všech ostatních lokálních identifikačních služeb (Local Handle Service), jež jsou zpětně propojené s globální centrální jednotkou. Dotaz ke směrování jakéhokoli identifikátoru proto může být identifikačnímu systému zaslán z libovolné služby.
3. **Každý identifikátor** (viz označení HS#n na obr. 2/3 pod číslem 3) může být spojen s jedním nebo více zapsanými údaji. Například již uvedený identifikátor *10.123/456* je spojen jak s URL, která zajišťuje komunikaci v rámci WWW, tak s obecným komunikačním repozitářovým přístupovým protokolem RAP (Repository Access Protocol), zajišťujícím komunikaci v rámci repozitáře. Taková informace je zpět poskytnuta po směrování klientovi (obr. 2/3, číslo 4). Propojit je možné i jeden identifikátor s vícenásobným digitálním objektem lokalizovaným na více URL.
4. V rámci procesů **administrativního klienta** dochází jednak k vytváření identifikátorů a doplňování a editování informací spojených s jejich směrováním (procesy označené číslem 5 na obr. 2/3), jednak k zasílání zpráv o úspěšných nebo chybových řešeních (procesy pod číslem 6 na obr. 2/3).

Aktuální informace o struktuře identifikátoru „handle“ (označovaný také zkratkou HDL) a jeho směrovacích službách, jakož i jiných identifikátorech digitálních objektů jsou uvedeny v 5. kapitole této práce

2.6.2 Digitální objekt

Digitální objekt (Digital Object) je v modelu „Kahn-Wilensky“ chápán v technickém slova smyslu jako specifický druh datové struktury [KAHN, 2006, s. 116-117], který zahrnuje (viz obr. 2/4):



Obr. 2/4: schéma částí digitálního objektu [ARMS, 1995b]

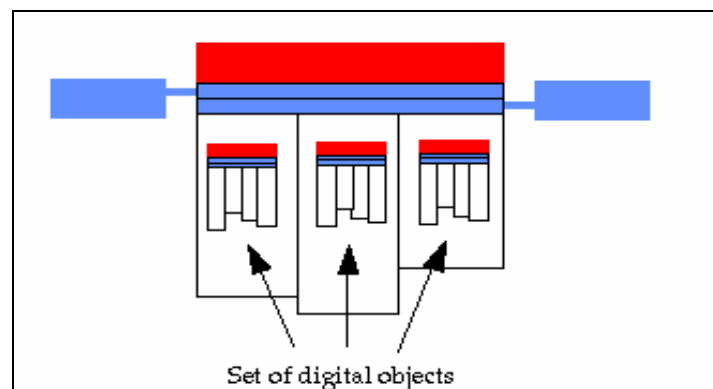
- jedinečný identifikátor (handle)
- obsah (content) jako sled bitů nebo soubor sledu bitů
- příslušné informace o vlastnostech (properties) digitálního objektu (metadata)
- příslušné informace o transakcích (transaction log), tj. o užití objektu jakékoliv povahy
- digitální signaturu (volitelná) garantující, že objekt nebyl změněn

Digitální objekty mohou být proměnlivé, pokud je dovoleno, aby byl jeho obsah měněn, nebo neproměnlivé.

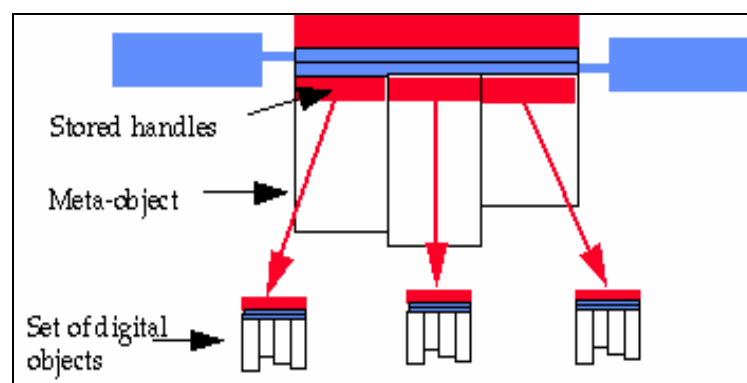
Digitální objekty mohou být jednoduché nebo může jít o soubor či skupinu (agregát) několika digitálních objektů.

2.6.3 Skupina (agregát) digitálních objektů

V rámci digitálních knihoven je v řadě případů nutné, aby jejich tvůrci seskupovali příbuzné digitální objekty. V repozitáři např. mohou být uloženy výzkumné zprávy ve více formátech (PostScript, PDF aj.) včetně jejich prezentací, různé interpretace hudebního díla aj. Skupiny digitálních objektů mohou být reprezentovány buď jako **složené objekty (composite objects)**, jejichž obsah zahrnuje soubor několika objektů (viz obr. 2/5), nebo jako **meta-objekty (meta-objects)**, jejichž obsah zahrnuje identifikátory pro soubor objektů - viz obr. 2/6.



Obr. 2/5: schéma složeného digitálního objektu [ARMS, 1995b]



Obr. 2/6: schéma digitálního objektu (meta-objektu) využívaného jako katalogizační záznam [ARMS, 1995b]

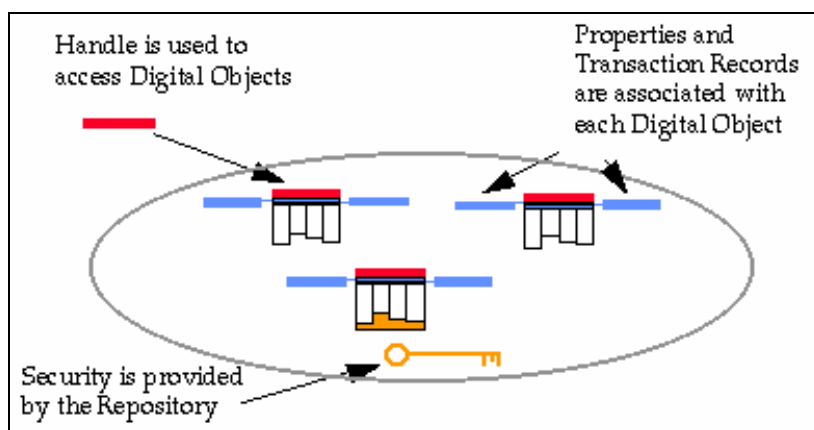
2.6.4 Digitální repozitáře (skladiště, úložiště)

Informace v digitální knihovně může být uložena v digitálních repozitářích (skladištích úložištích), která mohou mít formu:

- FTP archivu
- WWW serveru
- online databáze
- repozitáře (úložiště)

Repozitáře jsou vhodné pro materiály obsahující hodnotné informace. K základním charakteristikám repozitářů patří (viz obr. 2/7):

- v repozitáři jsou ukládány jak vlastní digitální objekty (strukturovaná data), tak příslušná metadata o jejich vlastnostech, struktuře, uložení apod. (z hlediska dnešní terminologie metadata popisná, strukturní, administrativní, technická) i informace o jejich užití (právní metadata)
- jeden objekt smí být uložen ve více repozitářích (replikace-kopie objektů)
- v repozitářích musí být zajištěna bezpečnost uložených objektů; individuální objekty mohou být (nepovinně) opatřeny signaturou, která může například garantovat platnost a neporušitelnost objektu apod.
- ke každému repozitáři náleží server s vyhledávacím systémem, který obsahuje informace o objektech v repozitáři; takový server nemusí být součástí repozitáře



Obr. 2/7: schéma repozitáře pro zabezpečené uložení digitálních objektů [ARMS, 1995b]

2.7 Referenční model Otevřeného archivního informačního systému (RM OAIS)

Popis modelu či rámce digitální knihovny „Kahn-Wilensky“, který v době svého vzniku ještě nezahrnoval otázky dlouhodobého uchovávání a ochrany digitálních objektů, je potřeba doplnit o popis dalšího obecného modelu či rámce digitální knihovny nebo digitálního archivu, který dostal název **Referenční model Otevřeného archivního informačního systému** (Reference Model for an Open Archival Information System, RM OAIS). RM OAIS je v současnosti již publikovanou mezinárodní normou [ISO 14721:2003]. Z hlediska obsahu je zaměřen na komunikaci strukturovaných informací s důrazem na proces jejich dlouhodobé archivace a ochrany a na proces jejich zpřístupňování uživatelům. Standard je již využíván

také v oblasti digitálních knihoven a elektronických archivů v oblasti vědy, výzkumu a vzdělávání, další projekty dlouhodobé archivace a ochrany těchto informací se chystají.

Dále uvedený text vychází z originální publikace modelu OAIS [Consultative Committee, 2002-2007], původní podrobné výukové prezentace modelu OAIS [SAWYER, 2002] a kvalitního sekundárního výkladu modelu OAIS [LAVOIE, 2004].

2.7.1 Vznik mezinárodní normy RM OAIS

Národní centrum pro vědecká data z oblasti vesmíru (National Space Science Data Center , NSSDC, <http://nssdc.gsfc.nasa.gov/>) amerického Národního úřadu pro aeronautiku a kosmický prostor (NASA, USA, <http://www.nasa.gov/>) je institucí, která se musela, kromě jiného, po desetiletí zabývat systematicky otázkami dlouhodobé archivace digitálních dat pocházejících z četných a nákladných vědeckých výzkumů a programů NASA. Na půdě NSSDC vznikl jeden z prvních digitálních archivů, který od roku 1966 prošel řadou technologických změn [SAWYER, 2002, snímek 3]. Získané zkušenosti centra NSSDC se nutně promítly také do práce **Konzultačního výboru pro systémy zpracovávající data z oblasti vesmíru** (Consultative Committee for Space Data Systems, CCSDS, <http://www.ccsds.org>), který je mezinárodním fórem agentur působících v oblasti kosmického výzkumu (vznikl v roce 1982). CCSDS vyvinul řadu nezávislých standardů pro zpracování digitálních dat, a i proto se stal v roce 1990 partnerem Mezinárodní standardizační organizace ISO, konkrétně její Technické komise TC 20 (pro oblast vzdušných a kosmických lodí) a zejména její subkomise SC 13, která se zabývá přípravou standardů pro digitální data z oblasti kosmického výzkumu a standardů pro systémy zajišťující přenos informací.

ISO navrhla, aby CCSDS připravil **standard pro archivaci digitálních dat** se zaměřením na data získaná z vesmírných misí. Zdůrazněn byl požadavek jejich **dlouhodobé archivace**. V té době neexistoval žádný jiný rámec pro rozvoj specifických standardů k digitálním archivům. V rámci subkomise SC 13 byla dohodnuta příprava **Referenčního modelu (RM)**, který měl ustanovit společné pojmy a termíny. Zároveň byla dohodnuta široká účast dalších zainteresovaných subjektů včetně tradičních archivů. Bylo také zdůrazněno, že standard se sice měl zaměřit na data v elektronické formě, ale zřetel měl být brán i na to, že ve většině tradičních archivů existují i jiné formy dat [SAWYER, 2002, snímek 5].

Referenční model (RM) měl být podle plánu subkomise SC 13 zcela **obecným rámcem** s cílem pochopit významné vztahy mezi **definovanými entitami** určitého prostředí a také rámcem pro vývoj **standardů** či **specifikací** takové prostředí podporující. RM měl být **abstrakcí** malého množství **klíčových pojmů**, jejich **vzájemných vztahů** i **vztahů k vnějšímu okolí**.

Z hlediska organizačního se vedení prací v subkomisi SC 13 ujali odborníci z NASA, kteří navázali spojení s odborníky z Federální komise pro geografická data USA (Federal Geographic Data Committee, FGDC, USA) a také odborníky z Národního úřadu USA pro archivy a záznamy (National Archives and Records Administration, NARA). Celý proces přípravy standardu měl být **otevřený** i pro odborníky z jiných komunit, průběžné výsledky přípravy standardu měly být prezentovány na konferencích a pracovních seminářích a publikovány na WWW. Vznikla e-mailová podpora projednávání textu připravovaného standardu.

Technické hledisko přípravy standardu vyžadovalo prozkoumání jiných referenčních modelů v té době již jako ISO standardy existujících (například známý sedmivrstevný Komunikační referenční model aj.). Měl být jasně definován pojem „**archivování dat**“ s požadavkem rozčlenit celý proces na **dílčí funkční oblasti: příjem, archivace a ochrana, přístup, plánování archivace** aj. Měl být definován soubor rozhraní mezi jednotlivými funkčními

oblastmi a také soubor kategorií dat pro potřeby archivace. Pro přípravu schémat toků dat ve funkčním modelu byl zvolen jazyk UML (Unified Modeling Language, ver. 1.1, 1997).

Referenční model je zaměřen na různé kategorie zájemců (projektanty archivů i jejich uživatele, správce archivů, vývojáře standardů v této oblasti aj.). Přijaté definice termínů jdou napříč různými oblastmi společenské praxe komunikace informací, ke kterým patří tradiční archivy, střediska vědeckých dat, digitální knihovny aj.

Referenční model OAIS byl přijat globálně jako důležité východisko pro dlouhodobou archivaci a ochranu digitálních dat. Uplatnil se již v době dokončování standardu v řadě projektů digitálních knihoven (v oblasti tradičních archivů, národních knihoven, středisek vědeckých dat a také komerčních organizací).

Subkomise SC 13 postoupila Referenční model OAIS jako návrh mezinárodní normy k veřejnému posouzení nejprve ve verzi z května 1997, později ve verzi z května 1999. Finální návrh mezinárodní normy byl definitivně schválen v červnu 2000. V lednu 2002 byla oficiálně publikována jako standard ISO 14721:2002. Editorské korekce na úrovni schémat byly publikovány v září 2007 [Consultative Committee, 2002-2007, s. v]. Text standardu je na základě dohody mezi ISO a CCSDS volně dostupný také na WWW [Consultative Committee, 2002-2007]. Standard má celkem 6 kapitol a 5 příloh (A-F).

V první úvodní části standardu RM OAIS je objasněn účel, rozsah, aplikovatelnost a principy modelu OAIS i směry budoucí práce. Jsou podány definice základních termínů. Druhá část představuje podstatné pojmy modelu OAIS, třetí pak odpovědnosti dílčích subsystémů. Čtvrtá část je věnována detailnímu představení funkčního a informačního modelu OAIS. Pátá část obsahuje informace o perspektivě archivace a ochrany dat a část šestá probírá otázky interoperability digitálních archivů.

2.7.2 Archiv typu OAIS jako ústřední pojem a jeho hlavní odpovědnosti (funkce)

Ústředním pojmem referenčního modelu je pojem „**otevřený archivní informační systém**“. Termín „**otevřený**“ se vztahuje, jak bylo výše v textu již uvedeno, jednak k procesu tvorby celého standardu, jednak k volnému zpřístupnění celého textu standardu různým odborným komunitám i širší veřejnosti. Termínem „**archivní informační systém**“ se v modelu rozumí organizace lidí a systémů (hardwaru a softwaru), která přijala odpovědnost za archivaci a ochranu informací a za jejich zpřístupňování určeným komunitám [Consultative Committee, 2002-2007, s. 1-1]. V této definici je zdůrazněn, jak upozorňuje B. F. Lavoie, proces **dlouhodobé archivace a ochrany informací** a proces **zpřístupňování archivovaných informací** způsobem, který bude vyhovovat **určeným komunitám uživatelů** [LAVOIE, 2004, s. 3]. Podstatným předmětem činnosti archivu typu OAIS jsou, jak vyplývá z výše uvedené definice, **informace**, které jsou z hlediska stanovených cílů archivního systému velmi přesně definovány, zejména pokud jde o jejich formy a typy (viz dále část 1.7.4).

Archiv typu OAIS se od jiných archivů liší definovanými **povinnými odpovědnostmi** (funkcemi). Patří k nim [LAVOIE, 2003, s. 4-5]:

- Dojednat a přijmout vhodné informace od jejich producentů; nutné je stanovit kritéria pro příjem vhodných informačních materiálů k archivaci (například předmětové vymezení, původ nebo formát materiálů); důležité je také motivovat producenty nebo majitele informačních materiálů k tomu, aby je chtěli svěřit do péče archivu
- Získat dostatečnou kontrolu nad informacemi z důvodu zajištění cílů jejich dlouhodobé archivace a ochrany; archiv by měl získat dostatečná intelektuální práva k tomu, aby mohl vedle základní správy archivovaných materiálů zajistit také nezbytné operace vyplývající

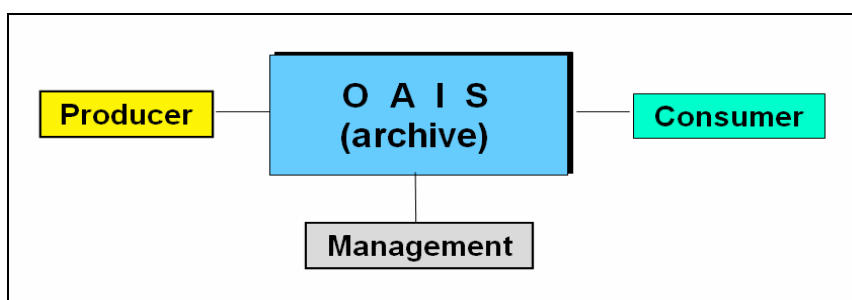
z komplexních cílů archivace (například pořizovat jejich nové verze, které si v budoucnu vyžádá případná změna informačních technologií aj.)

- Určit šíři základních (prvořadých) uživatelských komunit archivu; jde o podstatnou podmínku k výkonu dalších odpovědností archivu
- Zajistit, aby byly archivované informace nezávisle srozumitelné určeným uživatelským komunitám, to znamená, že archivovaným informacím by měli uživatelé porozumět zásadně bez pomoci jejich producenta; archiv typu OAI tudíž musí archivovat i dostatečné množství souvisejících kontextových informací (metadat), aby archivované informace byly srozumitelné a využitelné i budoucími generacemi uživatelů za sto nebo tisíc let; kontextové informace se mohou týkat například struktury nebo formátu archivovaného dokumentu, mohou zahrnovat vysvětlení, jak a kdy dokument vznikl, nebo jiné znalosti apod.)
- Dodržovat písemně stanovenou politiku a předepsané procedury archivu, aby bylo zajištěno, že informace budou v budoucnu uživatelům skutečně k dispozici, že budou chráněné před všemi nahodilými událostmi a že bude možné rozšiřovat autentické kopie archivovaných informací v jejich původní formě nebo formě odvozené z formy původní
- Zpřístupnit archivované informace pomocí vhodných mechanismů a služeb určeným uživatelským komunitám

2.7.3 Externí entity archivu OAIS a jejich odpovědnosti (funkce)

Referenční model OAIS popisuje a specifikuje jednotlivé definované entity, jejich funkce, odpovědnosti a procesy, které by měly v archivu probíhat. Model je aplikovatelný na jakékoliv typy archivů (tedy i archivy tradiční povahy), které mají za úkol dlouhodobě archivovat a chránit svěřené informace. Problém „**dlouhodobosti**“ je úzce spojen s budoucími změnami informačních technologií. Standard nespecifikuje žádnou konkrétní (softwarovou) implementaci archivního systému ani nestanovuje nějaké konkrétní druhy informačních objektů, které budou v systému archivovány a skrze archivní systém komunikovány směrem k uživatelským komunitám.

Okolí archivu, odpovídajícího požadavkům na referenční model OAIS, tvoří následující externí entity (viz obr. 2/8):



Obr. 2/8: schéma modelu okolí archivu typu OAIS (převzato a upraveno podle [SAWYER, 2002, snímek 15])

- **Producent (Producer)** – je rolí či funkcí osob (jednotlivců nebo korporací) nebo klientských systémů, které poskytují informace (informační objekty) k jejich dlouhodobé archivaci a ochraně; producenti předkládají informace společně s metadaty do archivu typu OAIS prostřednictvím procesu příjmu (angl. ingest process); vztahy mezi

producentem a archivem typu OAIS jsou formalizovány prostřednictvím stanovené (písemné) dohody (angl. agreement)

- **Management** – je rolí či funkcí těch, kteří stanovují a zavádějí celkovou politiku archivu typu OAIS jako jednu z komponent širší politiky v dané oblasti; management stanovuje strategické plánování archivace a ochrany informací, definuje záběr archivovaných informačních sbírek či fondů, stanovuje garance ochrany informací, může zajišťovat finanční zdroje pro potřeby archivu a také kontrolní funkce činnosti a výkonů archivu; management ale není odpovědný za správu každodenních operací archivu [LAVOIE, 2004, s. 5]
- **Spotřebitel (Consumer)** – je rolí nebo funkcí těch osob (jednotlivců nebo korporací) nebo klientských programů, jež komunikují se službami archivu typu OAIS tak, aby byly nalezeny a získány archivované informace, které jsou předmětem jejich zájmu; model OAIS definuje preferovanou specifickou skupinu spotřebitelů archivovaných informací označovanou termínem „**určená komunita**“ (angl. Designated Community); jde o spotřebitele (uživatele), od kterého se očekává, že bude schopen nezávisle porozumět archivované informaci v určité archivované formě (předpokládá se například, že archivovaný vědecký dokument nebo datový soubor z výzkumu bude moci v budoucnu bez další pomoci využívat uživatel, který bude mít odpovídající vzdělání a zkušenosti v daném vědeckém oboru); jiní uživatelé budou moci odbornou archivovanou informaci získat také, budou však pravděpodobně potřebovat další související metadatové či znalostní informace, aby ji porozuměli a mohli ji případně využít ve své činnosti; tvůrci modelu OAIS jinak také předpokládají případné dynamické změny v širší určené uživatelské komunitě v blízké či vzdálené budoucnosti

Všechny uvedené externí entity včetně samotného archivu typu OAIS představují spíše funkční než organizační pohled na jimi zajišťované role. Z toho vyplývá, že uvedené role mohou být realizovány buď na úrovni jediné instituce, nebo mohou být distribuovány na více institucí [LAVOIE, 2003, s. 7].

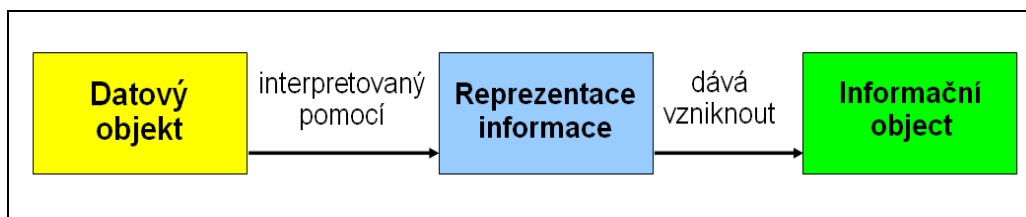
2.7.4 Informace, informační objekty a balíčky informací a jejich pojetí v modelu OAIS

Model OAIS je případem informačního systému, který má specifické poslání a specifické funkce. Jde-li o informační systém, pak to znamená, že jeho hlavním předmětem činnosti jsou informace a jejich komunikace. Model OAIS definuje větší množství toků (přenosů) informací mezi jednotlivými prvky celého systému. Závažnost rolí či funkcí modelu OAIS (výše popsaných) způsobila, že bylo nutné pojem „**informace**“ (information), resp. pojem „**informační objekt**“ (information object) vymežit, a to velmi přesně.

Informace je obecně v modelu OAIS (ale i v jiných informačních systémech) definována jako **jakýkoliv typ znalosti**, který může být komunikován (směňován). V průběhu komunikace je informace vždy reprezentována či vyjádřena pomocí dat [Consultative Committee, 2002-2007, s. 1-10] – informace je tedy abstrakcí, nehmotným sdělením reálných dat. Související pojem „**informační objekt**“ je v modelu OAIS podstatnou kategorií (obr. 2/9). Je definován jako kombinace **datového objektu** (obr. 2/9) a **reprezentace informace** (obr. 2/9) [Consultative Committee, 2002-2007, s. 1-11]. Schéma na obr. 2/9 lze doplnit ještě následujícím slovním popisem:

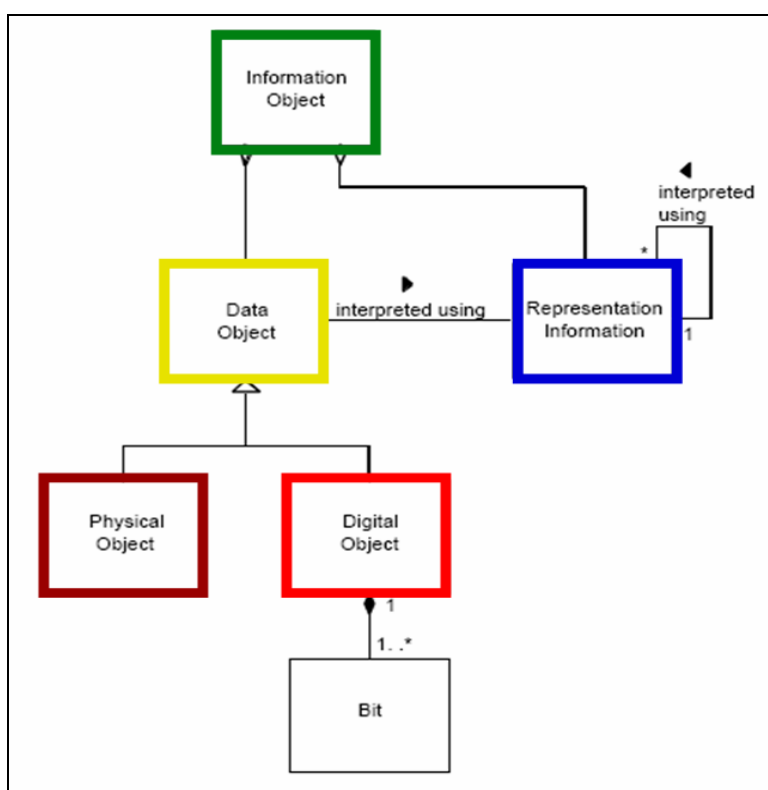
- Informace (informační objekt) je vždy vyjádřena (tj. reprezentována) pomocí nějakého typu dat

- Data (datový objekt) interpretovaná pomocí (svoji) reprezentace informace dávají vzniknout (nové) informaci (informačnímu objektu)



Obr. 2/9: schéma získání informace (informačního objektu) z dat (datového objektu) v modelu OAIS [Consultative Committee, 2002-2007, s. 2-4]

- Archivace informačního objektu v modelu OAIS vyžaduje jasnou identifikaci a pochopení datového objektu a také jeho související reprezentace informace

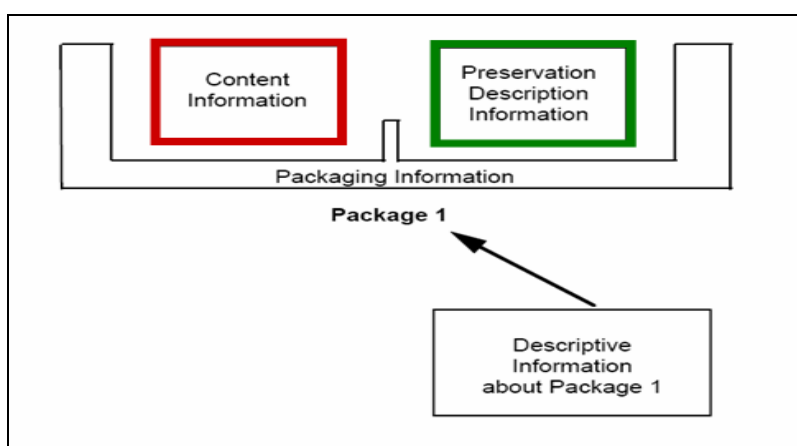


Obr. 2/10: schéma složení informačního objektu [Consultative Committee, 2002-2007, s. 4-20]

Informační objekt (Information Object, viz obr. 2/10) jako základní kategorie modelu OAIS se tedy, jak bylo výše uvedeno, skládá z **datového objektu** (Data Object), který může být buď fyzický (Physical Object) nebo digitální (sled bitů, Digital Object), a **reprezentace informace** (Representation Information), která umožňuje ucelenou interpretaci datového objektu do podoby smysluplné informace (informačního objektu).

Reprezentace informace datového objektu fyzické povahy (například zkameněliny trilobity) může (a nemusí) poskytovat doplňkový význam o tomto objektu (půjde o výsledek analýzy fyzicky pozorovatelných vlastností takového objektu). Taková informace může být připravena v průběhu času a může se pak stát součástí uceleného informačního objektu.

Reprezentace informace digitálního objektu (sled bitů) doprovází datový objekt, poskytuje o něm nutný doplňkový význam (například informace o typu digitálních dat, o skupinách či agregátech dat apod.). Řečeno jinými slovy, reprezentace informace digitálního objektu konvertuje sledy bitů do více smysluplné informace (informačního objektu) [Consultative Committee, 2002-2007, s. 4-21]. Může jít například o popis datového formátu. Model OAIS takový typ reprezentace informace označuje jako „**strukturní informaci**“ (Structure Information – viz obr. 2/12). Ta může být navíc doplněna o „**informaci sémantickou**“ (Semantic Information – viz obr. 2/12), která blíže strukturní informaci osvětluje (například může sdělovat význam hodnoty určitého prvku). Reprezentace informace digitálního objektu může také zahrnovat **odkaz na jinou reprezentaci informace**, jež může být jiným svébytným informačním objektem (například definovaný číselník kódů zemí). Na obr. 2/10 je taková situace znázorněna pomocí vazby ‘interpreted using’ (interpretovaná pomocí), uvedené u reprezentace informace. Může tak vzniknout celá „sít’ reprezentací“. Reprezentace informace má tedy v modelu OAIS **rekurzivní povahu** (může se neomezeně opakovat).



Obr. 2/11: schéma složení a vztahů balíčku informací v modelu OAIS (převzato a upraveno podle [Consultative Committee, 2002-2007, s. 2-5])

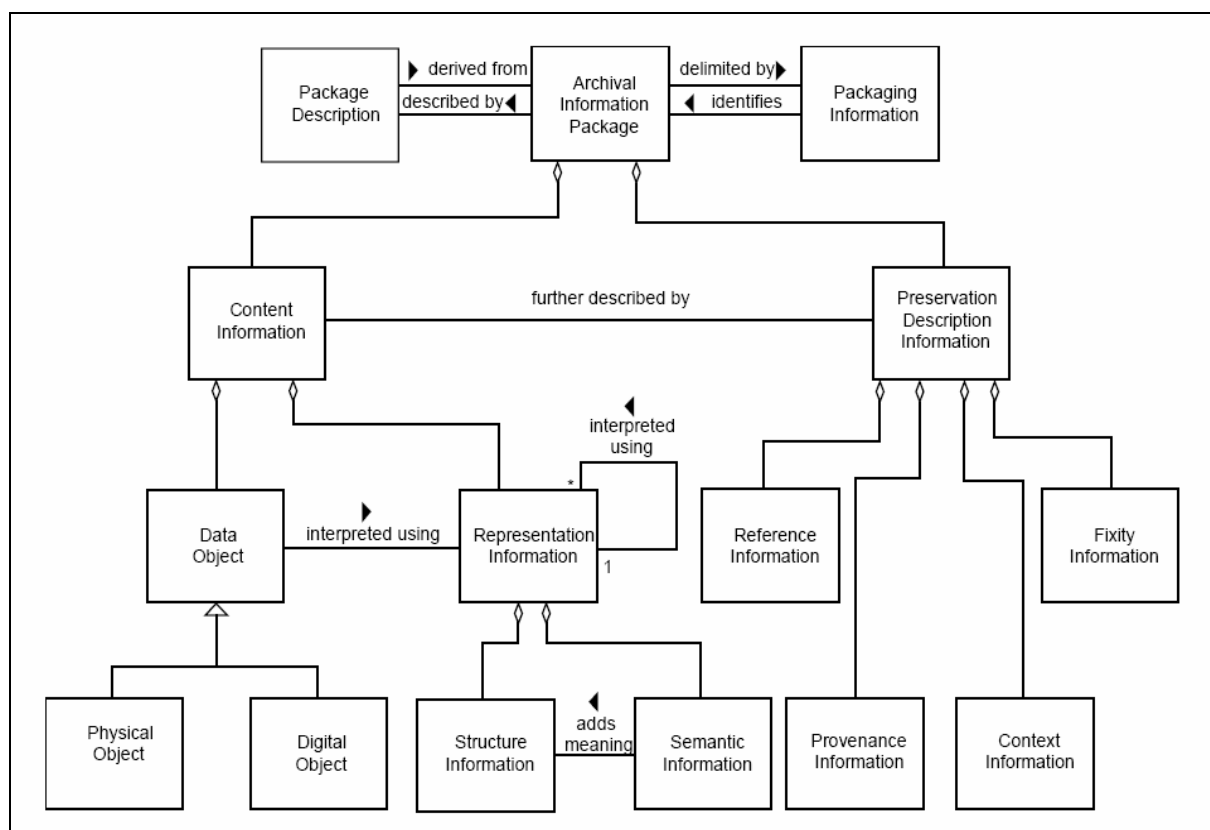
Pro potřeby komunikace informací (informačních objektů) vně a uvnitř archivu typu OAIS byl zaveden specifický termín „**balíček informací**“ (angl. information package; v češtině lze popřípadě uplatnit i širší termín „soubor informací“). Balíček informací je definován jako pojmový kontejner, který nese dva typy informací (viz obr. 2/11): vlastní **obsah informace** (angl. Content Information), který má být archivován a chráněn, a **archivační popisnou informaci** (Preservation Description Information, PDI), která má povahu souvisejících metadat podporujících zejména dlouhodobou archivaci. Metadata jsou modelem OAIS striktně požadována z důvodu zajištění cílů dlouhodobé archivace a ochrany. Celý balíček informací je opatřen další informací o zabalení obsahu informace a PDI do **jednoho logického balíčku** (Packaging Information). Jde o informaci, která identifikuje a lokalizuje tuto logickou jednotku v rámci archivního systému. Tzv. **deskriptivní informace** (angl. Descriptive information – viz obr. 2/11) je určena pro potřeby vyhledávání obsahu archivované informace určitou uživatelskou komunitou prostřednictvím vyhledávacích nástrojů archivu typu OAIS (může jít o metadatový záznam typu Dublin Core apod.).

Archivační popisná informace (Preservation Description Information, PDI) se skládá z následujících čtyř typů metadatových informací (viz obr. 2/12):

- **Referenční informace** (Reference Information) – jedinečným způsobem identifikuje obsah archivované informace jak uvnitř archivu OAIS, tak vně archivu (může jít o interní

identifikátor obsahu generovaný systémem, o trvalý mezinárodní identifikátor ISBN, ISSN, DOI, HDL, URN, ARK aj.)

- **Kontextová informace** (Context Information) – popisuje vztahy obsahu archivované informace k obsahům jiných informačních objektů (například vztah jednotky knižního titulu k edici apod.)
- **Informace o původu** (Provenance Information) – dokumentuje historii obsahu archivované informace (její vytvoření, jakékoli změny v obsahu nebo formátu vzniklé v průběhu času, realizované akce v archivu, jako například migrace dat apod.)
- **Informace o stabilitě** (Fixity Information) – zajišťuje autenticitu a integritu obsahu archivované informace (například pomocí digitální signatury, digitálního vodoznaku aj.)



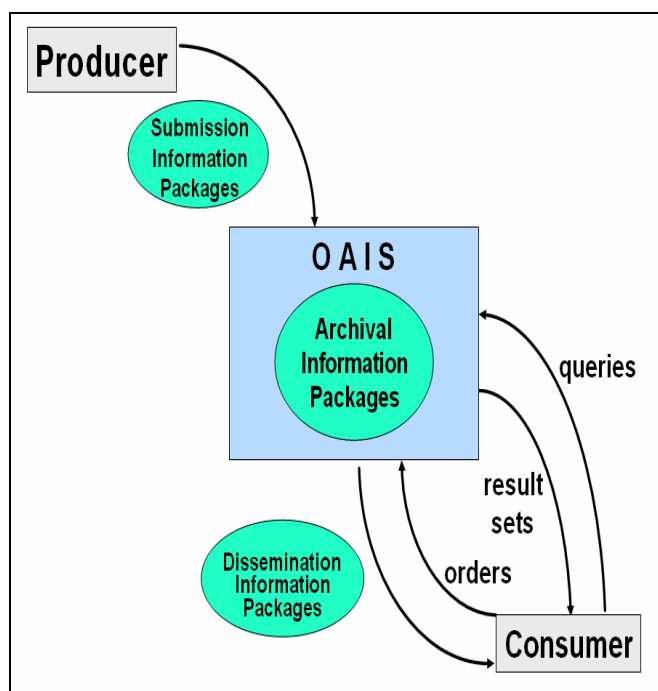
Obr. 2/12: schéma komplexního balíčku informací pro archivaci (AIP) (detailní pohled) (převzato z: [Consultative Committee, 2002-2007, s. 4-37])

V modelu OAIS jsou dále definovány 3 varianty balíčků informací [LAVOIE, 2003, s. 11]:

- **Balíček informací pro dodávání** (angl. Submission Information Package, **SIP**, viz obr. 2/13) – je verzí balíčku informací, který je předáván od producenta do archivu typu OAIS v rámci procesu příjmu; přesná forma balíčku SIP bývá zpravidla výsledkem smluvního ujednání archivu s producentem; pojem balíčku SIP zdůrazňuje fakt, že informace nemůže být archivována ve formě, ve které ji předal její producent (například pokud je dodána producentem ve formátu, který nepodporuje archiv OAIS, nebo pokud producent dodá neúplná metadata, jež musejí být následně v průběhu příjmu doplňována)
- **Balíček informací pro archivaci** (angl. Archival Information Package, **AIP**, viz obr. 2/13) – je verzí balíčku informací, který má být uložen a dlouhodobě archivován v archivu typu OAIS; balíček AIP se skládá z vlastního obsahu informace, který je hlavním

předmětem archivace a ochrany, a dále z kompletního souboru metadat, dostatečných k podpoře archivačních a vyhledávacích služeb archivního systému; obsah informace a metadata tvoří jeden logický balíček; systém uspořádání balíčku AIP pro potřeby archivace a ochrany si stanovují správci systému

- **Balíček informací pro rozšiřování** (angl. Dissemination Information Package, **DIP**, viz obr. 2/13) – je verzí balíčku informací, který je předáván či poskytován spotřebiteli (uživateli) archivu; pojem balíčku DIP zdůrazňuje fakt, že tento balíček se může co do formy a obsahu lišit od balíčku AIP (lišit se budou například formáty obsahu informace - například text uložený v archivu ve formátu XML bude uživateli poskytnut ve formátu PDF, obraz uložený ve formátu TIFF bude pro potřeby uživateli konvertován do formátu JPEG apod.); balíček DIP nemusí také obsahovat všechna archivovaná metadata



Obr. 2/13: schéma toku externích dat ve formě balíčků informací v modelu archivu OAIS (převzato a upraveno podle [Consultative Committee, 2002-2007, s. 2-8; SAWYER, 2002, snímek 20])

Hlavním předmětem zájmu archivu typu OAIS je balíček informací pro dlouhodobou archivaci a ochranu (AIP).

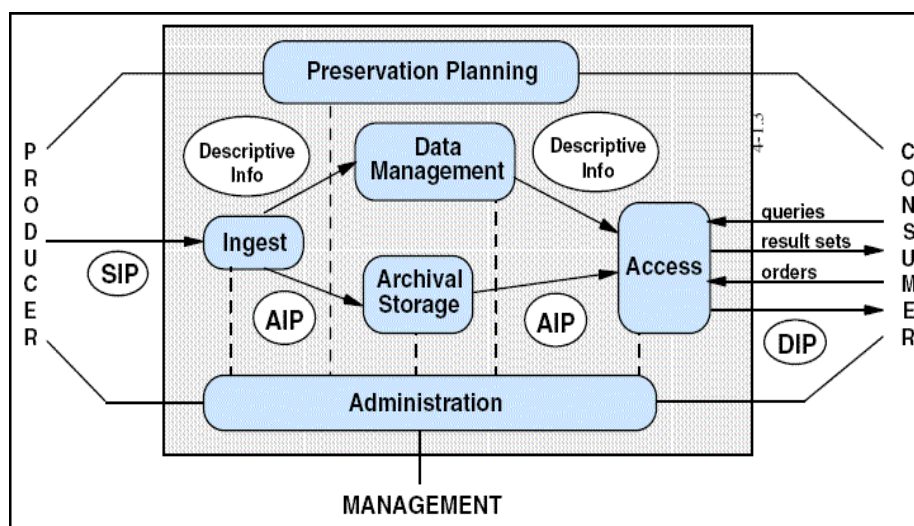
2.7.5 Funkční model archivu typu OAIS (základní úroveň)

Cílem přípravy funkčního modelu archivu OAIS bylo ukázat hlavní funkční oblasti archivace informačních materiálů. Jeho tvůrci využili metodu **funkční dekompozice**. Jejich záměrem bylo osvětlit řadu možných funkcionalit, se kterými se lze v archivu typu OAIS setkat. Cílem ale nebylo provést dekompozici všech možných úrovní činnosti archivu, tvůrci se zaměřili pouze a dvě: základní (ta je předmětem popisu této podkapitoly) a druhou. Další případné úrovně již nejsou zpracovány. Dekompozice měla zajistit užitečný soubor hlavních pojmů a termínů. Definované **dílní funkční komponenty (části či subsystémy)** druhé úrovně jsou relativně samostatnými celky, v modelu je jich zahrnuto celkem 6. Tvůrci modelu však předpokládali, že všechny archivy nebudou v praxi potřebovat implementovat všechny specifikované komponenty. Zároveň předpokládali, že v praxi budované archivy budou v moderních počítačových aplikacích zahrnovat řadu **společných podpůrných služeb**, jako

jsou služby pro zabezpečení komunikace mezi jednotlivými procesy, služby pro správu jmen, služby pro správu souborů (zálohování) a adresářů, služby pro zajištění bezpečnosti dat aj.) [SAWYER, 2002, snímek 40-41].

Funkční model zahrnuje následující komponenty (funkční části či subsystémy) [LAVOIE, 2004, s. 8-9] – viz také znázornění procesů na obr. 2/14:

1. **Příjem (angl. Ingest)** – je specifikován jako soubor procesů a služeb, odpovědných za příjem informací, tj. **balíčků informací pro dodávání (SIP)**, podaných jejich producenty, a za přípravu těchto informací k dlouhodobému uložení do repozitáře archivu (skladiště, úložiště). Ke specifickým funkcím příjmu patří: potvrzení příjmu informací přenesených do archivu; prověření (validace), že přijaté informace jsou kvalitní, kompletní a neporušené; transformace podaných informací do formy vhodné k uložení v archivu a k jejich další správě - jsou automaticky generovány **balíčky informací pro archivaci (AIP)**; extrakce (z balíčků AIP) a/nebo vytvoření **popisných metadat** (Descriptive info) k podpoře systémů či nástrojů u pro vyhledávání informací; přenos podaných informací (balíčků SIP) a s nimi souvisejících metadat k uložení v repozitáři archivu. Funkce příjmu slouží v archivu typu OAIS jako rozhraní mezi ním a externí entitou producenta.



Obr. 2/14: funkční schéma modelu OAIS základní úroveň (převzato z: [Consultative Committee, 2002-2007, s. 4-1])

2. **Archivní uložení (angl. Archival Storage)** – je druhou funkční komponentou (subsystémem) archivu typu OAIS. Jde o část vlastního archivního systému, který má za úkol správu dlouhodobé archivace a ochrany informací, tj. **balíčků informací pro archivaci (AIP)**, svěřených do péče archivu. Balíčky AIP jsou přebírány ze subsystému příjmu. Subsystém archivního uložení odpovídá za to, že archivovaný obsah zůstane ve vhodné formě uložení (např. online, off-line) a že sled bitů zahrnující archivovanou informaci zůstane kompletní a bude i za dlouhou dobu znovu interpretovatelný či srozumitelný. Funkce archivního uložení musí z důvodu cílů dlouhodobé archivace také průběžně provádět procesy, jako jsou **obnova médií**, **migrace dat** do nového formátu aj. Tento subsystém musí mít také implementované **záruční mechanismy** realizovaných procesů (automatické kontroly chyb, vyhodnocování výsledků archivačních procesů), musí mít připravenou **politiku obnovy** činnosti archivního systému pro případ havárií, **politiku pro zmírnění následků** mimořádných událostí aj. Archivní subsystém zajišťuje také dílčí operace vyhledání archivovaných informací (balíčků AIP), požadovaných

uživateli v rámci subsystému přístupu. Subsystém archivního uložení však sám o sobě nemá žádné přímé rozhraní pro komunikaci s externími entitami.

3. **Správa dat (Data Management)** – funkční subsystém zajišťující prioritně údržbu **databází popisných metadat** (Descriptive Info – viz obr. 2/14), která identifikují a popisují archivované informace (balíčky AIP), za účelem podpory vyhledávacích systémů či nástrojů archivu. Správa dat zabezpečuje také **administrativní metadata**, která podporují výkon operací interního systému (například statistiky přístupu aj.). K dalším funkcím tohoto subsystému patří **realizace dotazů** (queries) nad indexy metadatových databází, **generování zpráv** na základě požadavků jiných funkčních subsystémů celého archivu. Zabezpečuje rovněž **aktualizaci databází**, jakmile jsou do archivu zařazeny nové materiály, nebo poté, co jsou z určitých důvodů nějak modifikovány, nebo jsou z archivu zcela vyřazeny.
4. **Administrace (Administration)** – je funkční subsystém odpovědný za **správu všech každodenních operací** celého archivního systému. Koordinuje aktivity ostatních pěti subsystémů. Zajišťuje také interakce s producenty informací (sjednávání dohod aj.), se spotřebiteli (zajišťování podpory služeb) a správou dat (zavádění a údržba archivní politiky a standardů). Administrace je také zodpovědná za **dohled** nad operacemi archivace a systémy přístupu, **monitoruje výkony** celého systému a koordinuje aktualizace archivního systému.
5. **Plánování archivace a ochrany (Preservation Planning)** – služba odpovědná za podrobné **plánování strategie archivace a ochrany** informačních materiálů uložených v archivu OAIS a také za doporučování **nezbytných revizí této strategie** v reakci na vývoj podmínek v okolí archivu OAIS. Tato služba **monitoruje** změny v externím prostředí, které mohou mít vliv na schopnost archivu chránit a udržovat přístup k informacím, jako jsou **inovace** v procesech ukládání informací a **inovace** v technologiích přístupu, nebo změny v požadavcích určených uživatelských komunit. Subsystém plánování archivace připravuje **doporučení pro aktualizaci politiky a procedur** archivu, aby bylo možné se změnám přizpůsobit. Plánování archivace znamená záruku archivu proti neustále se měnícím požadavkům uživatelů a měnícím se okolním technologiím. Pomáhá při implementaci strategie archivace v samotném archivním systému. Přípravuje **plány digitální migrace** (je definována jako přenos digitální informace za účelem archivovat ji v archivu typu OAIS). Stimulem k migraci digitálních dat (týká se především balíčků informací AIP) může být rozpad počítačových nosičů, požadavky na zlepšení efektivity nákladů na činnost archivu nebo nové požadavky na uživatelské služby [Consultative Committee, 2002-2007, s. 5-1-5-2].
6. **Přístup (Access)** – poslední funkční subsystém spravující procesy a služby, pomocí kterých spotřebitelé, v první řadě určené komunity, lokalizují, žádají a získávají informace, uložené v repozitáři archivu. Typickou službou subsystému přístupu je **zpracování dotazu** (angl. query) uživatele na existující fondy archivu. Požadavky jsou předány ke zpracování do subsystému správy dat, jenž po realizaci dotazu nad metadatovou databází předkládá odpovědi zpět v podobě souboru metadatových záznamů (results sets), který je poté předán uživateli. V další fázi subsystém přístupu koordinuje operace vyhledání a dodání konkrétního požadovaného obsahu uloženého v repozitáři archivu (balíčku AIP). Požadavek (order) je předán k vyřízení do subsystému archivního uložení, po získání balíčků AIP jsou v subsystému přístupu realizovány nezbytné transformační operace s předávaným obsahem (například reformátování dat) dříve než je definitivně předán uživateli **balíček informací pro rozšiřování (DIP)**. Subsystém přístupu představuje rozhraní s externí entitou spotřebitele (uživatele).

Referenční model OAIS detailněji (na druhé úrovni) specifikuje (v části 4.1.1) všech šest výše jmenovaných funkčních subsystémů [Consultative Committee, 2002-2007, s. 4-3–4-16]. Komplexní složené detailní funkční schéma celého archivního systému zahrnuje původní text standardu zcela na konci textu v příloze F [Consultative Committee, 2002-2007, s. F1].

2.7.6 Referenční model OAIS a praxe

Referenční model OAIS je materiálem výlučně **koncepční povahy**. Nezahrnuje žádné pokyny k jeho implementaci či nasazení v reálné praxi. Neposkytuje informace o architektuře archivního systému, o doporučených technologiích, které by měly zajišťovat zpracování dat a jejich uložení, o databázovém systému ani o žádných dalších technických detailech k zajištění navržených funkcí systému. RM OAIS je jen možným (ale žádoucím) východiskem pro budování archivních informačních systémů, zejména digitální povahy.

Již v době tvoření a diskutování textu standardu RM OAIS využila řada institucí a iniciativ tento model k zahájení prací v oblasti dlouhodobé archivace informačních materiálů. Byly navrženy první **systémy repozitářů** odpovídajících požadavkům RM OAIS, první **standarty** podporující funkce OAIS, byly zpřesněny požadavky na metadata potřebná pro dlouhodobou archivaci, byly vyvinuty metody a protokoly pro kódování a komunikaci archivovaných informací [LAVOIE, 2004, s. 14]. Z četných aktivit lze jmenovat například následující:

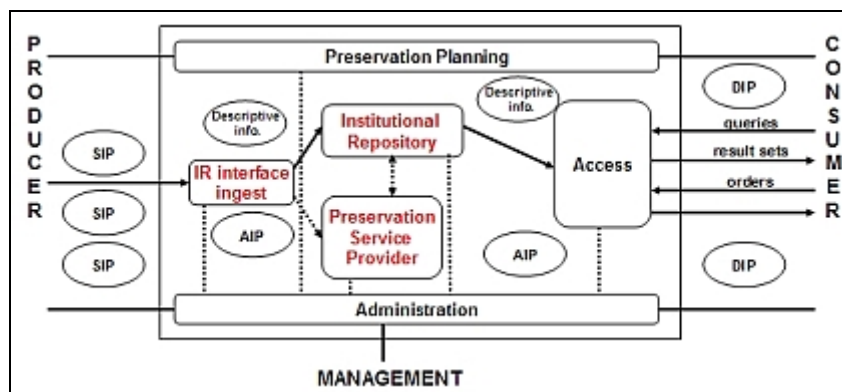
- Již výše zmíněné Národní centrum NSSDC (NASA) rozvíjí svůj digitální archiv pomocí RM OAIS jako základu nové architektury
- Národní archiv a správa archivních záznamů USA podepsala dohodu o spolupráci v oblasti dlouhodobé archivace svých fondů a sbírek se Superpočítačovým centrem v San Diegu - bylo ohlášeno využití RM OAIS
- Národní australská knihovna využila RM OAIS taktéž jako základu svých informačních implementací
- Významný evropský projekt *NEDLIB* (Networked European Deposit Library) využil RM OAIS jako základu designu a architektury dílčího systému DSEP (Deposit System for Electronic Publications) pro potřeby dlouhodobé archivace elektronických publikací
- Královská knihovna v Nizozemsku a později také Německá národní knihovna nasadily do svého systému digitální knihovny nákladný systém (software) *DIAS* (Digital Information Archiving System) firmy IBM, který je založen na modelu OAIS
- Pojmy modelu OAIS využívá i jeden z nových metadatových standardů *METS* (Metadata Encoding and Transmission Standard), který je udržován a rozvíjen na serveru Kongresové knihovny
- Problematiku dlouhodobé archivace dokumentů sledovaly v rámci výzkumných aktivit také hlavní americké sítě – bývalá RLG (Research Libraries Group) a OCLC (Online Computer Library Center) – po jejich spojení v roce 2006 sleduje intenzivně tyto úkoly již jen OCLC; podstatné jsou dva směry výzkumu: 1. v oblasti metadat, která mají zajistit dlouhodobou archivaci digitálních materiálů – úsilí vyústilo do návrhu standardu pro archivační metadata *PREMIS* (Preservation Metadata: Implementation Strategies), 2. studium vlastností spolehlivých digitálních repozitářů, příslušné doporučení bylo zveřejněno formou zprávy poprvé v roce 2002 (Trusted Digital Repositories : Attributes and Responsibilities), později v únoru 2007 (Trustworthy Repositories Audit & Certification : Criteria and Checklist (PURL), <http://bibpurl.oclc.org/web/16712>); v evropském kontextu byl pro potřeby kontroly spolehlivosti repozitářů vyvinut nástroj DRAMBORA (<http://www.repositoryaudit.eu/>)

2.7.6.1 Modely využití RM OAIS v oblasti institucionálních repozitářů

Otázka dlouhodobé archivace informačních materiálů z oblasti vědy a výzkumu zajímá dnes již také řadu provozovatelů institucionálních nebo mezinárodních centralizovaných repozitářů. Až dosud se tvůrci a vývojáři softwarových aplikací, zabezpečujících provoz digitálních archivů, soustřeďovali spíše na otázky jejich základních funkcí (například funkce vyhledávání informací, vstupu dat aj.). Otázky zajištění a/nebo podpory dlouhodobé archivace zůstávaly zatím více méně stranou. Situace se ale zlepšuje, tlak narůstajícího množství digitálních materiálů a potřeba jejich ochrany zesiluje. Dva z nejužívanějších softwarů provenience USA - **DSPACE** (<http://www.dspace.org>) a **FEDORA** (<http://www.fedora.info/>) již ve svých posledních verzích podporu dlouhodobé archivace zajišťují. Známa britská softwarová aplikace **EPrints** (<http://www.eprints.org/>), která je čteně využívána zejména ve Velké Británii, ale i jinde ve světě, prozatím nabízí tuto podporu méně zřetelně. Faktem však zůstává, že prakticky žádný z existujících institucionálních akademických repozitářů není schopen sám o sobě zajišťovat v tuto chvíli náročné úlohy dlouhodobé archivace obsahu svých repozitářů [HITCHCOCK, 2007, Introduction], tak jak je definuje model OAIS.

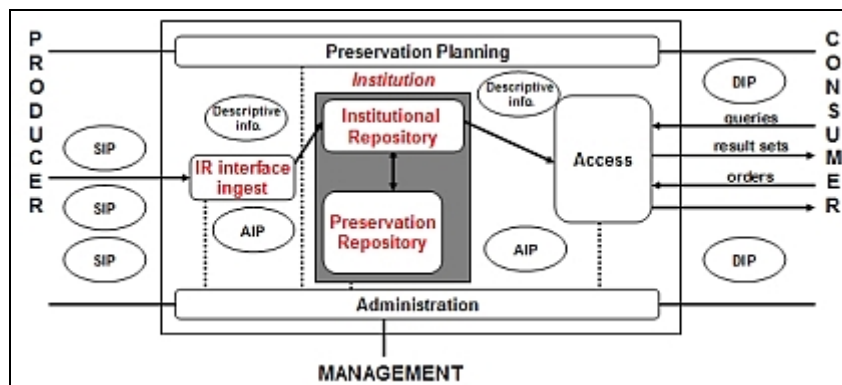
Skupina odborníků Southamptonské univerzity realizovala v nedávné době projekt **Preserv** (<http://preserv.eprints.org>), zkoumající a testující možnosti dlouhodobé archivace digitálního obsahu z cca 200 britských akademických repozitářů. V publikovaném článku z června 2007 autoři, kromě jiného, představili možné modely zajištění specifických funkcí dlouhodobé archivace obsahu akademických repozitářů s důrazem na skutečnost, že existující softwarové aplikace samy o sobě zajišťují běžně řadu funkcí definovaných modelem OAIS (vstup dat, vyhledávání metadatových informací, poskytování služeb aj.) [HITCHCOCK, 2007, Three OAIS preservation models for IRs]. Na základě obecného modelu RM OAIS představili tři modely zajištění specifických archivačních služeb:

- **Model provozovatele služby** (Service provider model, viz obr. 2/15), který předpokládá, že podstatnou funkci OAIS, tj. archivního uložení (Archival Storage) bude zajišťovat externí provozovatel služby (Preservation Service Provider), například národní archiv, digitální úložiště šedé literatury, národní knihovna apod., ostatní funkce si bude zajišťovat institucionální repozitář sám včetně funkce příjmu informací (Ingest – IR interface ingest) a správy dat (Institutional Repository); tento model byl v rámci projektu Preserv otestován (ve spolupráci s britským Národním archivem (TNA) a také Britskou knihovnou); s externím provozovatelem archivních služeb (Královská knihovna) se počítá také v národním nizozemském systému pro vědu a výzkum NARCIS, který v rámci svých služeb využívá zejména digitální obsahy všech nizozemských univerzitních repozitářů

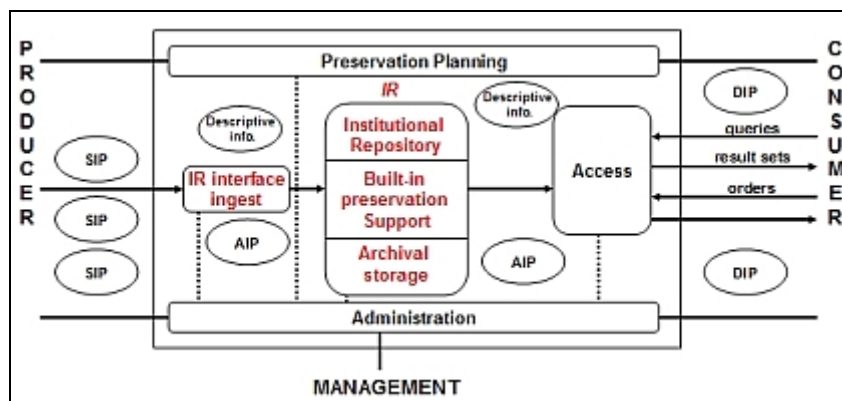


Obr. 2/15: model zajištění dlouhodobé archivace institucionálního repozitáře externím provozovatelem služby (převzato z: [HITCHCOCK, 2007, obr. 2b])

- **Model instituce** (Institutional model, viz obr. 2/16), který předpokládá, že důležité funkce OAIS, tj. funkci správy dat (Data Management – Institutional Repository) a funkci archivního uložení (Archival Storage – Preservation repository) bude zajišťovat instituce, v jejímž rámci je provozován daný repozitář; repozitářů může být i více, mohou být provozovány například v softwaru EPrints, Fedora aj.; ostatní služby si budou zajišťovat jednotlivé institucionální repozitáře



Obr. 2/16: model zajištění dlouhodobé archivace institucionálního repozitáře institucí, která provozuje repozitář (převzato z: [HITCHCOCK, 2007, obr. 2c])



Obr. 2/17: model zajištění dlouhodobé archivace samotným institucionálním repozitářem (převzato z: [HITCHCOCK, 2007, obr. 2d])

- **Model institucionálního repozitáře** (Institutional repository (IR) model, viz obr. 2/17), který předpokládá, že všechny funkce OAIS bude zajišťovat sama softwarová aplikace; půjde také o funkce správy dat (Data Management – Institutional Repository) a funkci archivního uložení (Built-in preservation Support / Archival Storage)

3 Otevřený přístup

Narůstající fondy informačních materiálů z oblasti vědy a výzkumu, které jsou dostupné zdarma v prostoru WWW, jsou předmětem zájmu nejen **koncových uživatelů**, kteří je stále více vyhledávají a využívají ve své práci, ale také předmětem značného zájmu informačních profesionálů a řady **automatizovaných informačních služeb**, které tyto materiály zpracovávají a dále zpřístupňují ve formě zajímavých výstupů. Ke službám značného významu v oblasti vědy a výzkumu patří nejen četné **vyhledávací služby** (nad více elektronickými archivy nebo digitálními knihovnami), ale například také služby citační, které, kromě jiného umožňují sledovat vliv vědeckých prací na další vývoj vědeckého výzkumu. Text této kapitoly představuje problematiku informačních zdrojů, které jsou šířeny v prostředí WWW zdarma. Definován je termín „**otevřený přístup**“, představeny jsou **hlavní typy jeho zdrojů** s volně dostupnými informačními materiály, jejich **mezinárodní registry** a také **instituce a iniciativy**, které otevřený přístup k informacím podporují.

3.1 Definice otevřeného přístupu

Komunikace informací ve vědě doznala v průběhu posledních 18 let velkých změn. Výrazným trendem dneška je další rozšiřování otevřeného přístupu k informacím. Termín „**otevřený přístup**“ (angl. Open Access) v sobě zahrnuje význam **trvalého a bezplatného online přístupu k dokumentům, zejména úplným textům, pro všechny uživatele**. (fyzické, korporativní i „strojové“). Otevřený přístup se vyznačuje dvěma podstatnými rysy: dokumenty jsou **dostupné bezplatně**, i když nemusejí být bezplatně vytvářeny, a vlastník copyrightu dává jednoznačné svolení k jejich neomezenému **čtení, stahování, kopírování, sdílení, ukládání, tištění, vyhledávání a hypertextovému propojování** [HARDY, 2005]. Lze doplnit, že otevřený přístup je v rámci tohoto příspěvku míněn vždy v kontextu bezplatného zpřístupňování informací v prostředí WWW. Termín „otevřený přístup“ lze tedy ztotožnit s termínem „**bezplatný online přístup**“. V různých oborech lze bezplatně poskytovat různé typy dokumentů (preprinty, články z recenzovaných časopisů, technické zprávy, konferenční sborníky aj.).

Prvním významným reprezentantem bezplatného online přístupu k informacím ve vědě byl mezinárodní archiv elektronických preprintů vědeckých recenzovaných článků **arXiv.org** (<http://arXiv.org/>), založený komunitou fyziků již v roce 1991 (nyní je provozován na Cornellově univerzitě v Ithace, USA). Vznik WWW na počátku 90. let 20. stol. a jeho rychlé rozšiřování po celém světě vedlo řadu autorů i z dalších oborů k vystavování kopií publikovaných vědeckých článků volně (více méně neorganizovaně) na jejich **personálních webových stránkách**, lokalizovaných zpravidla na univerzitních serverech. To nebyl optimální jev, proto začaly na přelomu tisíciletí podle modelu archivu arXiv.org postupně vznikat institucionální repozitáře elektronických tisků, do kterých autoři mohli pomocí procesu **auto-archivace** svoje dokumenty uložit a zároveň popsat. Velký impuls k tomu dala **Iniciativa otevřených archivů** (Open Access Initiative, OAI, <http://www.openarchives.org/>), která se zrodila v roce 1999 právě za účelem integrace lokálních repozitářů prostřednictvím sklizení jejich metadat na základě protokolu **OAI-PMH** (Protocol for Metadata Harvesting) [LAGOZE, 2008]. Koncepce OAI však počítala se všemi typy dokumentů, které se mohou v digitálních knihovnách vyskytovat. Lze konstatovat, že se tak z koncepce OAI do značné míry vytratil původní cíl bezplatného online přístupu jen k recenzovaným časopisům [HARDY, 2005].

Tato myšlenka opět ožila a byla formalizována jako „otevřený přístup“ v rámci tzv. **Budapešťské iniciativy otevřeného přístupu** (Budapest Open Access Initiative, BOAI,

<http://www.soros.org/openaccess/read.shtml>). Zdůrazněny byly dva způsoby otevřeného přístupu: uložit preprint článku do univerzitního nebo tematického repozitáře a pokusit se potom předložit tento text buď do komerčního nebo otevřeného online časopisu.

3.2 Významné události před nástupem hnutí otevřeného přístupu (do roku 1991)

Tato část textu čerpá z přehledové práce známého experta v oblasti otevřeného přístupu prof. Petera Subera (Earlham College, USA, <http://www.earlham.edu/~peters/>). Je nazvaná *Časový přehled hnutí otevřeného přístupu* [SUBER, 2008b] a autor v ní stručnou formou průběžně řadu let zaznamenává podstatné události související s otevřeným přístupem s důrazem na oblast vědy, výzkumu a vzdělávání.

Vznik hnutí otevřeného přístupu lze v souladu s P. Suberem, vymezit zhruba rokem 1990, nicméně již před tímto datem lze zaznamenat řadu událostí (především pak z regionu USA), které současnou intenzivní vlnu tohoto hnutí ve světě předznamenaly či ji ovlivnily. Autor zmiňuje založení dvou mezinárodních databází v roce 1966 – ERIC (<http://www.eric.ed.gov/>) a MEDLINE (http://www.nlm.nih.gov/databases/databases_medline.html), které jsou dnes díky principům informační politiky USA zcela volně dostupné celému světu. V roce 1970 k nim přibyla také databáze AGRICOLA (<http://agricola.nal.usda.gov/>). Tyto databáze produkují významné národní knihovny USA. V roce 1974 byla na Stanfordské univerzitě založena databáze pro komunitu fyziků SPIRES (<http://www.slac.stanford.edu/spires/>), která je také volně dostupná uživatelům z celého světa. Kromě jiného zajišťuje citační službu pro archiv arXiv.org. Na konci srpna 1969 byl Ministerstvem obrany USA zahájen provoz bývalé sítě ARPANET (Advanced Research Projects Agency Network) – její činnost byla ukončena v roce 1990.

Významným počinem pro provoz později vzniklé sítě Internet bylo založení dnes významného volně dostupného digitálního fondu textů internetových standardů označovaných zkratkou RFC (Request for Comments, <http://www.rfc-editor.org/>, ISSN 2070-1721). V rámci sídla Editoru RFC je k dispozici již objemná databáze (<http://www.rfc-editor.org/rfc.html>), která kromě jiného, podává svědectví o vývoji této světové sítě. První text s pořadovým číslem RFC 1 byl zveřejněn již 7. dubna 1969 (<http://www.rfc-editor.org/rfc/rfc1.txt>). Dnes již známá volně dostupná rozsáhlá digitální knihovna elektronických knižních titulů *Projektů Gutenberg* (<http://www.gutenberg.org/>) zahájila činnost v roce 1971 díky Michaeli Hartovi.

Pro další vývoj otevřeného přístupu v oblasti vědy a výzkumu v USA bylo důležité zveřejnění Nařízení č. 189 (z 21. 9. 1985) prezidenta R. Reagana k otázce *Národní politiky přenosu vědeckých, technických a inženýrských informací* (<http://www.fas.org/irp/offdocs/nsdd/nsdd-189.htm>). Sdělovalo, kromě jiného, aby byl co největší objem produktů pocházejících z výzkumu dostupný veřejnosti bez omezení. V roce 1987 byla v USA založena významná iniciativa v oblasti elektronického zpracování úplných textů TEI (Text Encoding Initiative, <http://www.tei-c.org/>), v jejímž rámci se zrodila také značkovací jazyk i formát stejného jména (včetně metadatového popisu). Počínaje rokem 1989 nastalo období zakládání **volně dostupných online časopisů**. Patří k nim i recenzovaný mezinárodní časopis (založený v roce 1989 jiným neúnavným zastáncem volného přístupu prof. Stevanem Harnadem), který dostal název *Psychology* (<http://psycprints.ecs.soton.ac.uk>). Časopis je sponzorován Americkou psychologickou společností. Ve stejném roce založil Ch. W. Bailey významný časopis *The Public-Access Computer Systems Review* (PACS Review) pro obor počítačové vědy (<http://eprint.lib.uh.edu/pr/pacsrev.html>). Po nich následovala série dalších vědeckých časopisů, které přecházely na platformu volného přístupu.

V říjnu 1990 (dnes slavný) prof. Tim Berners-Lee naprogramoval prvního webového klienta a webový server. Zveřejnění této podstatné události se konalo v březnu 1991. 12. listopadu 1991 jmenovaný autor zveřejnil text s názvem *WorldWideWeb : návrh na HyperTextový projekt* (<http://www.w3.org/Proposal>) a o den později napsal první webovou stránku. Stal se tak průkopníkem a **otcem či zakladatelem služby WWW**. 17. května 1991 T. Berners-Lee společně s organizací CERN (Conseil Européen pour la recherche nucléaire, <http://public.web.cern.ch/public/>) zveřejnil první WWW standard (v roce 2002 CERN zveřejnil a zdarma poskytl software pro zabezpečení provozu elektronických archivů CDSWare (<http://cdsware.cern.ch/>) a v roce 2006 ohlásil projekt retrospektivní digitalizace všech článků z oblasti fyziky, které má ve svém fondu a jejich volné zpřístupnění).

3.3 Hlavní zdroje s otevřeným přístupem k informacím (vývoj volného přístupu po roce 1991)

K nejdůležitějším typům informačních zdrojů s otevřeným přístupem v současné patří:

- a) předmětově profilované mezinárodní archivy (centralizované nebo decentralizované)
- b) tzv. „institucionální repozitáře“ (dále také jako IR) (univerzitní, akademické aj.)
- c) online časopisy s otevřeným přístupem (vědecké, odborné)
- d) povrchový web (personální weby autorů, korporativní sídla aj.)

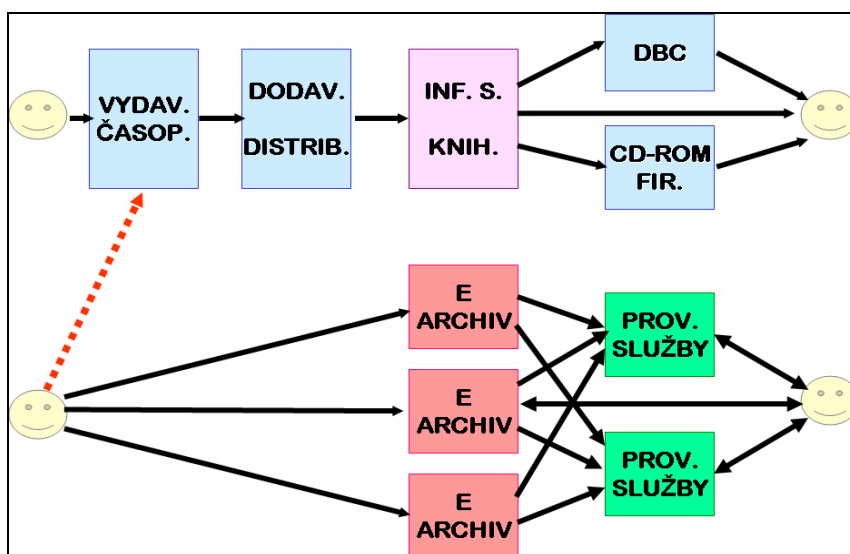
3.3.1 Mezinárodní elektronické archivy a digitální knihovny

16. srpen 1991 je památným datem v hnutí otevřeného přístupu. Komunita fyziků v čele s prof. Paulem Ginspargem se po kritické analýze klasického toku vědeckých poznatků od autora k uživateli (viz obr. 3/1), který pohlcoval značné objemy finančních prostředků a hlavně znamenal časové ztráty při komunikaci informací mezi vědci, rozhodla tuto situaci radikálně změnit. V rámci informačního systému v Národní laboratoři v Los Alamos (USA) byl založen jeden z nejvýznamnějších elektronických archivů s otevřeným přístupem – preprintový archiv *arXiv.org* (<http://arxiv.org>) pro několik exaktních oborů, především pro fyziku. Znamenal novou a hlavně rychlou cestu přímé komunikace vědeckých znalostí mezi autory a uživateli prostřednictvím volně dostupných archivů elektronických tisků (preprintů článkových informací). Autoři poskytují svoje texty volně do archivu s časovým předstihem s tím, že pokud chtějí mohou text později poskytnout i do recenzovaného komerčního časopisu (viz obr. 3/1). Provozování archivu *arXiv.org* má v současné době ve své péči Cornellova univerzita (USA) s podporou Národní nadace pro vědu (National Science Foundation, NSF, USA).

Archiv **arXiv.org** (<http://arxiv.org/>) je reprezentantem typu mezinárodních centralizovaných **předmětově profilovaných archivů**. Je „jedničkou“ z více hledisek v této kategorii, zahrnuje preprinty a postprinty vědeckých článků, ale také jinou šedou literaturu z oboru fyziky, matematiky, počítačové vědy, nelineárních věd, kvantitativní biologie a dalších nově zařazených specifických disciplín. Jde o jeden z nejstarších a také jeden z největších archivů (v registru ROAR zaujímá co do počtu dokumentů pátou příčku), jeho repozitář ke konci roku 2008 obsahoval již kolem 515 000 elektronických dokumentů (3. 10. 2008 bylo dosaženo počtu 500 000). Takových archivů je v současnosti řádově **několik desítek**. Registrační systém DOAR (o něm viz dále v textu) má v této souvislosti zavedenou kategorii typu repozitáře „**Disciplinary**“ (Oborové repozitáře) – tuto kategorii má také výše zmíněný arXiv.org.

V roce 1992 byl zahájen (dnes již historický) projekt virtuální digitální knihovny digitalizovaných textů z oboru počítačové vědy z různých amerických univerzit nazvaný

CSTR (Computer Science Technical Reports, <http://www.cnri.reston.va.us/home/cstr.html>), který trval až do roku 1996. Kromě jiného řešil řadu podstatných otázek **reálného provozu síťově dostupné digitální knihovny**. Jeho výsledkem byl také **obecný model digitální knihovny** (viz byly uvedeny již v části 1.6 tohoto textu).



Obr. 3/1: schéma klasické (v horní polovině obrázku) a novodobé (v dolní polovině obrázku) volné a otevřené komunikace vědeckých dokumentů (preprintů) od jejich tvůrců k jejich uživatelům

Legenda: Vydav. Časop. (vydavatelé komerčních časopisů), Dodav. Distib. (dodavatelé a distributoři informací), Inf. s. Knih. (informační střediska, knihovny), DBC (datázová centra), CD-ROM fir. (firmy produkující a zpřístupňující informace na CD-ROM, popř. DVD nosičích), E archiv (elektronické archivy) a Prov. služby (provozovatelé informačních služeb, zejména poskytování metadat)

Od roku 1993 začaly být zakládány další elektronické archivy s technickými zprávami. V tomto roce americká NASA (<http://www.nasa.gov>) uvedla do provozu Langleyský server technických zpráv (The Langley Technical Report Server). Thomas Krichel v únoru 1993 založil distribuovanou digitální knihovnu nazvanou *WoPEc* (Working Papers in Economics, dnes část zvaná EconPapers, <http://econpapers.repec.org/>), která se později (v květnu 1997) stala součástí dnes rozsáhlé distribuované digitální knihovny pro oblast ekonomických studií *RePEc* (<http://repec.org/>). V listopadu 1993 uvedl CERN, poté co jeho vedení odsouhlasilo bezplatné využívání celé aplikace včetně softwaru, do provozu svůj dokumentační preprintový server (<http://cdsweb.cern.ch/>), který patří k největším elektronickým archivům v kategorii „institucionální repozitáře“. V roce 1994 byl v USA spuštěn národní program USA s názvem *Iniciativa pro digitální knihovny – fáze I*. (Digital Libraries Initiative – Phase I), který později v roce 1998 vstoupil do druhé fáze (<http://www.dli2.nsf.gov/>). Jeho úkolem byla koordinace aktivit a projektů v oblasti digitálních knihoven na významných amerických univerzitách. Ve stejném roce (1994) se v návaznosti na původní systém CSTR zrodila také známá světová virtuální knihovna NCSTRL pro šedou literaturu z oboru počítačové vědy (Networked Computer Science Technical Reference Library, <http://www.ncstrl.org/>), založená původně na federativním modelu vyhledávání informací, od roku 2001 využívala již služeb sklizených metadat ze systému ARC (<http://arc.cs.odu.edu/>, vznikla v roce 2000). Hlavní stránka systému NCSTRL je ještě k dispozici, vyhledávání již ale není možné, neboť data byla převedena do jiného systému – popis tohoto systému je

v kapitole 4). Rok 1995 je rokem zrodu významné iniciativy pro popis elektronických zdrojů DCMI a související standardy Dublin Core (<http://dublincore.org/>).

V roce 1996 byla Virginským polytechnickým institutem a Státní univerzitou (Virginia Polytechnic Institute and State University) založena a zprovozněna první verze mezinárodní Sítově dostupné digitální knihovny disertačních a diplomových prací (Networked Digital Library of Theses and Dissertations, NDLTD, <http://www.ndltd.org/>), jejíž činnost, standardy a jiná doporučení znamenaly průlom v budování digitálních archivů vysokoškolských kvalifikačních prací v globálním rámci. Ve stejném roce byl zahájen provoz Internetového archivu (<http://www.archive.org/index.php>). 26. června roku 1997 byl americkým Národním centrem pro biotechnologické informace spuštěn provoz komplexního systému PubMed s tím, že jeho významná součást - báze MEDLINE, jak bylo výše již uvedeno, se stala volně dostupnou. Již výše zmíněný prof. Stevan Harnad založil ve stejném roce elektronický archiv *CogPrints* (<http://cogprints.org>), zaměřený na psychologii, neurovědu, lingvistiku a některé oblasti počítačové vědy (detaily viz kap. 4 v této práci). V roce 1997 byl Výzkumným institutem nadnárodní společnosti NEC spuštěn prototyp významného systému s názvem *CiteSeer* (po jistou dobu se nazýval *ResearchIndex*), který je svým charakterem digitální knihovnou vědecké literatury s citačními službami pro obor počítačové vědy a nově také informační vědy (<http://citeseer.ist.psu.edu/cs>) [BRATKOVÁ, 2001]. Data systému koncem roku 2007 migrovala do nové moderní databázové a webové aplikace s názvem *CiteSeerX* (<http://citeseerx.ist.psu.edu/>). V září 1998 byl z iniciativy reprezentantů archivu arXiv.org, virtuální knihovny NCSTRL a amerického vydavatelství ACM projekt společného volně dostupného systému *CoRR* (<http://arxiv.org/corr/home>, viz kapitola 4).

Většina z mezinárodních centralizovaných repozitářů není tak vysoko objemová co do počtu dokumentů, zejména pokud jde o sociální nebo humanitní obory. Výjimkou je pouze centralizovaný elektronický archiv postprintů **recenzovaných článků PubMed Central** (<http://pubmedcentral.nih.gov/>) pro biomedicínské obory (1 507 000 dokumentů, je na první pozici co do objemu dat v registru ROAR), a dále již výše zmíněná digitální knihovna s citačním rejstříkem **CiteSeerX** (<http://citeseerx.ist.psu.edu/>), která ve své nové verzi zahrnuje již téměř 1 170 000 zdrojových dokumentů (a kolem 23 000 000 citací). Archiv pro společenské vědy spravovaný ve Velké Británii **CogPrints** (<http://cogprints.org/>) pro kognitivní vědy (psychologii, neurovědu aj.) má jen cca 3 250 dokumentů a archiv **E-LIS** (<http://eprints.rclis.org/>) pro knihovní a informační vědu aj. jen cca 8 666 dokumentů. Provozy mezinárodních centralizovaných repozitářů zajišťují zpravidla určité univerzitní nebo vědeckovýzkumné organizace z nějaké země. Mohou využívat i případnou finanční podporu dalších korporací včetně mezinárodních. K nejvýznamnějším **mezinárodním archivům decentralizovaného (distribuovaného) typu** patří již zmíněný rozsáhlý systém pro disertační a diplomové práce **NDLTD** (<http://www.ndltd.org/>), jehož metadatové záznamy mohou libovolné informační služby sklízet přes protokol OAI-PMH také z databáze služby XTCat NDLTD (<http://alcme.oclc.org/ndltd/>), která zahrnuje již téměř 880 000 záznamů. Dalším objemným decentralizovaným systémem je systém **RePEC** (<http://www.repec.org/>), zajišťující ekonomické obory s typickou rozsáhlou skladbou informačních služeb (zahrnuje v současnosti ke konci 2008 celkem cca 761 000 dokumentů).

Rok 1999 je významný založením mezinárodní virtuální *Iniciativy otevřených archivů* (Open Archives Initiative, OAI, <http://www.openarchives.org>). Jejím hlavním produktem je protokol pro sběr metadat z otevřených archivů OAI-PMH (nyní již ve verzi 2.0 z roku 2002), který vznikl na bázi původní (z 22. října 1999) *Konvence ze Santa Fe* (http://www.openarchives.org/sfc/sfc_entry.htm). Ve Velké Británii byl ve stejné době založen čtyřletý projekt systému *OpCit* (Open Citation Project, <http://opcit.eprints.org/>) na podporu rozvoje citačních služeb nad volně dostupnými elektronickými archivy. Navázal na

něj od 10. 12. 2001 systém citační služby *CiteBase Search* (<http://www.citebase.org/>), kterou založil a rozvíjí Tim Brody. 26. 4. 1999 britské komerční vydavatelství *BioMed Central* (<http://www.biomedcentral.com/>) ohlásilo plán nabídky volného online přístupu ke svým časopisům (v červenci 2000 toto vydavatelství zveřejnilo první volně dostupný online článek). Od ledna 2002 vydavatelství zahájilo pravidelné účtování poplatků autorům na krytí nákladů na bezplatný přístup pro uživatele z celého světa. Od roku 2000 pak podobné služby v USA zajišťuje systém *PubMed Central* (<http://www.pubmedcentral.nih.gov/>). V USA ve stejném, roce založeno další neziskové vydavatelství vydavatelství recenzovaných vědeckých časopisů *PLoS* (Public Library of Science, <http://www.plos.org>), které na základě finanční podpory Moorovy nadace začalo od prosince 2003 volně zpřístupňova texty článků z časopisů.

The screenshot displays the ROAR website interface. At the top, there's a navigation bar with links like Home, Browse, Search, Content Search, Register a Repository, Help, Login, and Register in ROARMAP. Below this, search filters are visible: 'Any Geographic Country', 'Any System Software', and 'Any Content Type'. A dropdown menu for 'Any System Software' is open, listing various software types and their counts, such as 'Other softwares (various) (345)', 'ARNO (5)', 'Bepress (77)', 'CDS Invenio (13)', 'DigTool (4)', 'DIVA (16)', 'DoKS (5)', 'DSpace (373)', 'EDOC (2)', 'EPrints (286)', 'ETD-db (25)', 'Fedora (12)', 'Fez/Fedora (5)', 'Greenstone (2)', 'HAL (10)', 'HAL-Tor (1)', 'MyCoRe (4)', 'Open Journal System (4)', 'Open Repository (12)', 'OPUS (Open Publications System) (29)', and 'SciX (2)'. The main content area shows search results for 'Repositories (ROAR) resulted in 1232 repositories'. It includes a section for '@RCHIVESIC (994 records)' with a small thumbnail and a list of records. Below that, there's a detailed entry for 'Running HAL, based in France and is registered as Research Cross-Institutional'. This entry includes registration details (Registered on 2002-05-17), cumulative deposits (994 total), daily deposits in the last year (51 days of 1-9, 0 days of 10-99, 0 days of 100+), and links to various tools like [table], [graph], [interactive graph], [OAI Interface], [Identify List Metadata Formats List], and [harvest status]. Another entry for 'Running DSpace, based in United Kingdom and is registered as Research Institutional or Departmental' is also visible, showing registration on 2006-01-27, 218 total cumulative deposits, and 28 days of 1-9, 1 day of 10-99, 0 days of 100+ in the last year. The bottom of the screenshot shows a browser status bar with the URL 'http://roar.eprints.org/ind' and a 100% zoom level.

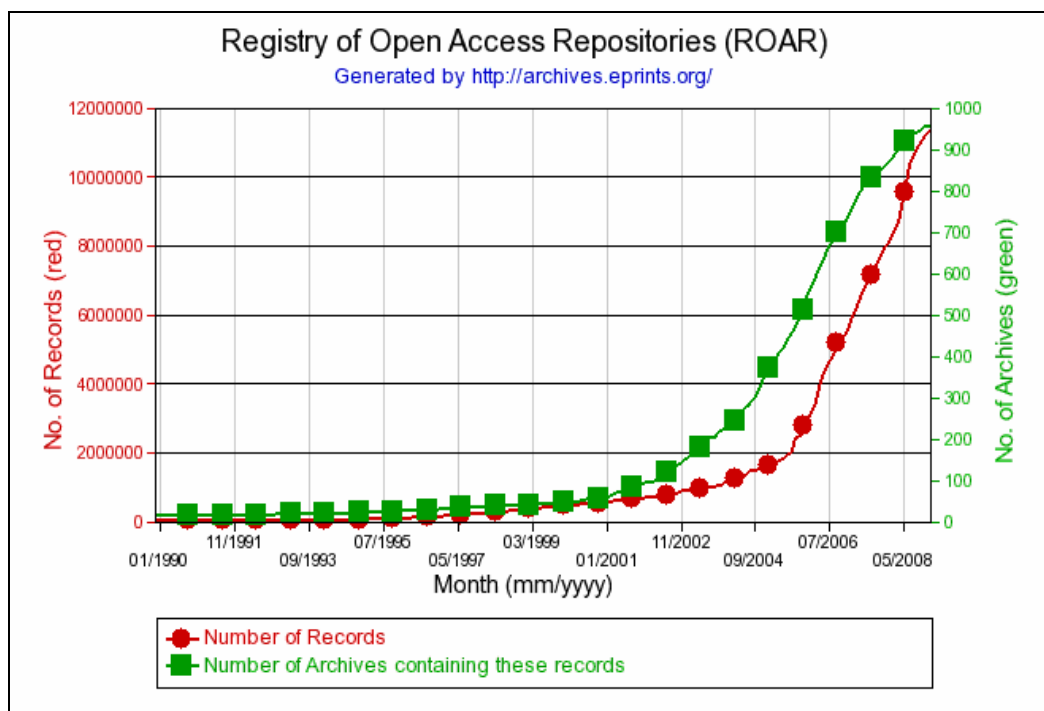
Obr. 3/2: hlavní rozhraní pro vyhledávání informací z registru ROAR s částí nabídky softwarů pro elektronické archivy [získáno 2008-12-15]

3.3.2 Institucionální repozitáře

Druhým typem archivů jsou **institucionální repozitáře** (Institutional repositories, **IR**), které začaly vznikat po roce 2000 na univerzitách nebo jiným akademických pracovištích po celém světě. K největším IR patří **Archiv Evropského sdružení pro jaderný výzkum** s téměř 100 000 dokumentů (úplných textů), systém však v základu zpřístupňuje cca 950 000 bibliografických záznamů vědecké literatury - k prosinci 2008 (CERN Document Server, <http://cdsweb.cern.ch>), a archiv **ECS** s více než 11 500 dokumenty (ECS EPrints Service, <http://eprints.ecs.soton.ac.uk>), který spravuje Fakulta elektroniky a počítačové vědy Southamptonské univerzity. Hlavním úkolem IR je zajistit otevřený přístup k recenzovaným výsledkům výzkumu dané instituce a jejích autorů, zároveň tím ale také zajistit větší vliv

jejich prací a v důsledku toho zvětšit také viditelnost a prestiž dané instituce. S nasazováním nových informačních technologií se objevují nové **externí citační služby**, jež pomáhají vliv a prestiž zvyšovat. Objevují se studie, které dokladují tezi, že dokumenty dostupné v otevřeném přístupu získávají značný vliv (impakt) ve srovnání s dokumenty dostupnými komerční cestou [HARNAD, 2004].

Informační vstupy do archivů/repozitářů jsou zpravidla zajišťovány pomocí procesu **auto-archivace**. Graf tempa růstu počtu institucionálních (včetně dalších typů archivů) archivů a počtu dokumentů k 22. 12. 2008 je zachycen na obr. 3/3. Návštěvnost největších mezinárodních archivů koncovými uživateli (vědci) z celého světa je enormní (desítky tisíc spojení denně).



Obr. 3/3: graf růstu institucionálních repozitářů (zelené body) a jejich digitálních obsahů (červené body, generovaný z registru databáze ROAR k 22. 12. 2008 (zajišťuje služba Celestial)

Elektronické archivy mohou být z hlediska programového zabezpečení provozovány buď v komerčních nebo volně dostupných softwarech. K nejrozšířenějším volně dostupným patří americký program **DSpace** (<http://www.dspace.org/>), vyvinutý na Massachusettském technologickém institutu (MIT, USA) – má k 16. 12 2008 podle registru ROAR již 373 instalací (viz. obr. č. 3/2), a také britský program **EPrints.org** (<http://software.eprints.org/>), který je rozvíjen na Southamptonské univerzitě – má podle registru ROAR již 286 instalací po celém světě. Oba programy jsou již připraveny plně na sklizení metadat podle protokolu OAI-PMH libovolnými informačními službami. V roce 2003 pak byla Virginskou univerzitou ve spolupráci s Cornellovou univerzitou světa nabídnuta další softwarová aplikace pro elektronické archivy – program **FEDORA** (<http://www.fedora.info/>).

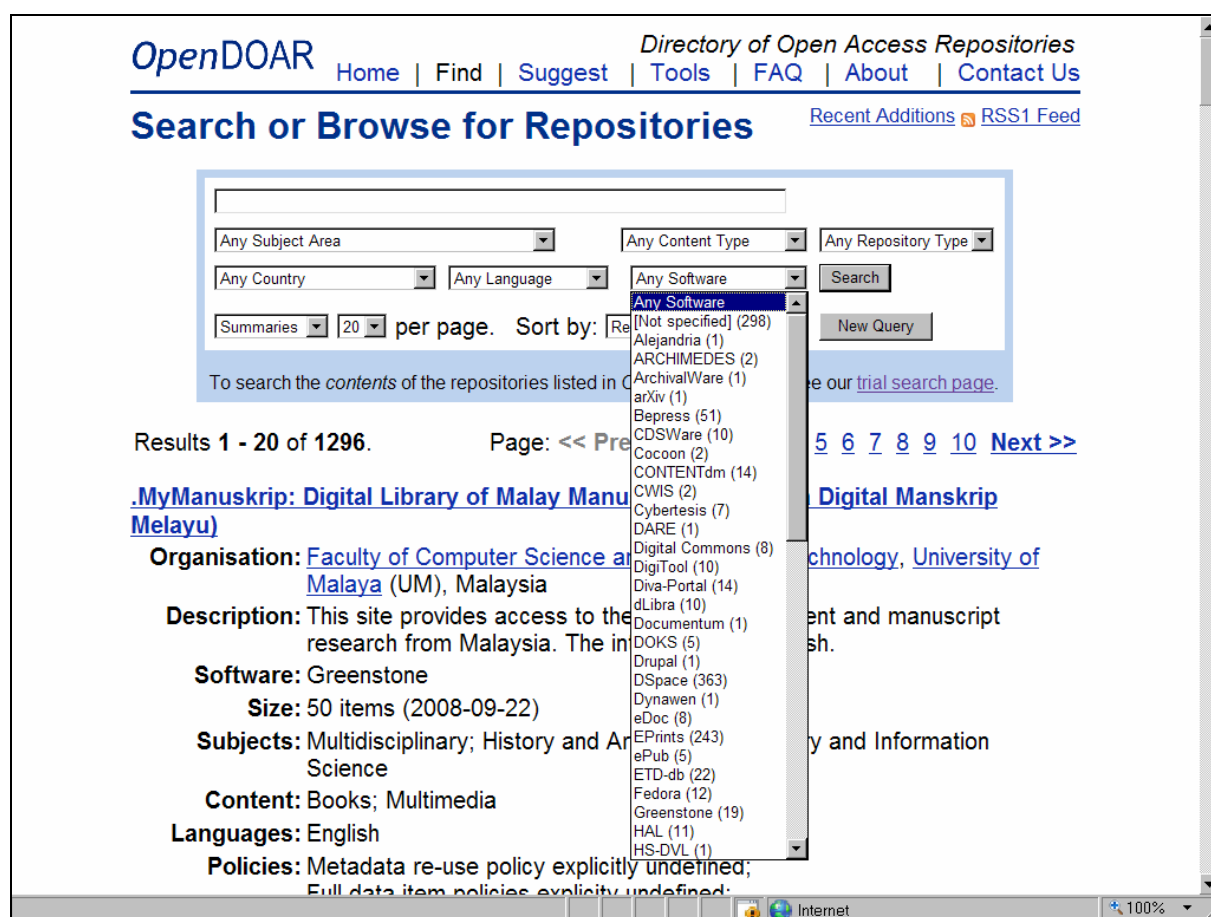
K významným službám, které indexují metadata z volně dostupných archivů a zajišťují efektivní vyhledávání informací nad repozitáři akademických archivů, je například služba **OAster** (<http://www.oaister.org/>), která vznikla v roce 2002 (zajišťovaná Knihovny Michiganské univerzity, USA).

Zajímavou službou je i bezplatný vyhledávací systém **Scirus** (<http://www.scirus.com/srsapp/>) komerčního vydavatele Elsevier, který kromě dat ze svých publikací zpracovává také data z volně dostupných institucionálních repozitářů, nikoliv ale ze všech.

3.3.3 Registry elektronických archivů

Předmětové profilované mezinárodní archivy a institucionální repozitáře jsou předmětem registrace již dvou databází:

1. Světová registrační databáze **ROAR** (Registry of Open Access Repositories, <http://roar.eprints.org/>) je spravována odborným týmem Southamptonské university (University of Southampton) pod vedením Tima Brodyho. K 16. prosinci 2008 bylo v databázi registrováno **1 232 repozitářů** různého typu. Databázová aplikace umožňuje z hlavní stránky pohodlné prohlížení záznamů repozitářů na základě 3 filtrů (podle země, aplikovaného softwaru a typu zahrnutých dokumentů). K dispozici je také prohlížení statistik (Browse) – viz jeden graf na obr. 3/3, vyhledávání záznamů pomocí klíčových slov (Search) a také fulltextové vyhledávání realizované pomocí systému „Google Custom Search“ (Content Search). Specifické rozhraní umožňuje nahlášení zájmu o registraci nového repozitáře. Záznamy repozitářů zahrnují zajímavé statistiky včetně grafů a obrazové náhledy hlavních stránek systémů repozitářů (viz obr. č. 3/2).



Obr. 3/4: hlavní rozhraní pro vyhledávání informací z registru DOAR s částí nabídky softwarů pro elektronické archivy [získáno 2008-12-15]

2. Světová registrační databáze **OpenDOAR** (Directory of Open Access Repositories, <http://www.opendoar.org/>) je spravována odborným týmem Nottinghamské university

(University of Nottingham) pod vedením Stephena Pinfielda K 16. prosinci 2008 bylo v databázi zaregistrováno téměř **1 300 repozitářů**. Systém registru umožňuje pohodlné vícehlediskové **vyhledávání záznamů repozitářů** (Search or Browse for Repositories) v režimu prohlížení i v režimu dotazového vyhledávání pomocí řady specifických hledisek – též podle typu repozitáře a využití softwarové aplikace (viz seznam hodnot nabytých v rámci rešeršního rozhraní na obr. 3/4). Nabízen je specifický rejstřík vyhledávání záznamů repozitářů **podle zemí a institucí** (OpenDOAR - Countries and Organisations). Na základě kvalitně vedené databázové registrace je pro uživatele automaticky generována řada statistických výstupů

Registr OpenDOAR nabízí na svých stránkách ještě také vyhledávání nad plnotextovým obsahem archivů či repozitářů (Search Repository Contents) - služba je zajišťována ve spolupráci se společností Google a její vyhledávací službou „Custom Search Engine“, (<http://www.google.com/coop/cse/>).

3.3.4 Volně dostupné online časopisy

První polovina 90. let 20. století i celé následující období je pro otevřený přístup významná také zakládáním volně dostupných vědeckých časopisů včetně časopisů recenzovaných. V roce 1990 byl založen recenzovaný online časopis *The Electronic Journal of Communication* (<http://www.cios.org/www/ejcmmain.htm>), v roce 1991 časopis *Ejournal* věnovaný otázkám elektronických sítí (<http://www.ucalgary.ca/ejournal/>) a také časopis *Behavioral and Brain Science* (<http://www.bbsonline.org>), který již zahrnoval také FTP preprintový archiv s otevřeným přístupem, v roce 1994 recenzované časopisy *Electronic Green Journal* (<http://egj.lib.uidaho.edu/index.php/egj>) a *Electronic Journal of Sociology* (<http://www.sociology.org/>), v roce 1995 známý časopis pro oblast digitálních knihoven *D-Lib magazine* (<http://www.dlib.org/>) nebo významný časopis v oblasti informační vědy, založený prof. T. D. Wilsonem *Information Research* (<http://informationr.net/ir/>) aj. Řada redakčních rad vědeckých komerčních časopisů začala v průběhu let vydávat „prohlášení o nezávislosti“, ve kterých protestovali proti dosavadní vydavatelské politice svých vydavatelů s doporučením zakládání alternativních online časopisů ve prospěch otevřeného přístupu. Tyto deklarace jsou sledovány koalicí SPARC (<http://www.arl.org/sparc/DI/>). Koalice pro vědecké publikování a akademické zdroje SPARC (<http://www.arl.org/sparc/>) vznikla na podporu otevřeného přístupu v roce 1998, je mezinárodní aliancí akademických a výzkumných knihoven (iniciovala ji americká Asociace výzkumných knihoven). Ve stejném roce prof. S. Harnad založil diskusi Otevřené fórum pro americké vědce o otevřeném přístupu (<http://american-scientist-open-access-forum.amsci.org/archives/American-Scientist-Open-Access-Forum.html>).

Volně dostupné online časopisy, které mohou zachovávat tradiční charakteristiky klasických časopisů, jsou určeny k bezplatnému využívání koncovými uživateli. Jejich významným rysem zpravidla bývá budování archivu či repozitáře k ukládání všech publikovaných článků. Někteří vydavatelé zajišťují volný přístup ke všem článkům (například již zmíněný vydavatel časopisů Public Library of Science, <http://www.plos.org/>). Jejich provoz je finančně zajišťován sponzory, případně i internetovou reklamou. Někteří nakladatelé praktikují zpřístupňování nových článků až po uplynutí určité doby (6 či více měsíců), starší články z archivu bývají dostupné bezplatně. Komerční nakladatelé mohou zveřejňovat bezplatně články z recenzovaných časopisů s tím, že náklady hradí autor nebo instituce za více autorů v částkách kolem 1 200-1 500 euro z ajeden článek (například BioMed Central, <http://www.biomedcentral.com>, který publikuje asi 100 online časopisů s otevřeným přístupem). Mnohé online časopisy jsou připraveny na sklizeň metadat na základě protokolu

OAI-PMH, takže se dostávají, stejně jako je tomu u elektronických archivů/repozitářů, ke zpracování v nadstavbových bibliografických a citačních službách.

Různé registry volně dostupných časopisů vykazují v současné době kolem **3 800 titulů** (v prosinci 2008, převážná část z nich je z oborů přírodních a aplikovaných věd), což je již solidní číslo, ale zatím značně menší ve srovnání s přibližně **24 000 tituly** recenzovaných komerčních časopisů [HARDY, 2005]. Registraci online volně dostupných časopisů zajišťuje s velkou péčí nejvýznamnější registrační systém **DOAJ** (Directory of Open Access Journals, <http://www.doaj.org/>), který spravuje Systém knihoven Lundské univerzity (Švédsko). V rámci registru je od června 2004 k dispozici také online **vyhledávání na úrovni článků z 1 320 online časopisů (237 430 záznamů článků)**.

3.3.5 Volně dostupné vědecké a jiné dokumenty na WWW

Elektronické kopie vědeckých článků z recenzovaných časopisů nebo i libovolné jiné odborné texty řada autorů vystavuje na svých **personálních webových stránkách**. Pro koncové uživatele se stávají vyhledatelnými jedine prostřednictvím univerzálních webových vyhledávacích služeb nebo prostřednictvím specifických citačních služeb, jako jsou **Google Scholar** nebo **CiteSeer**, resp. jeho nová verze **CiteSeerX** (Beta verze). Jde o zdroje neorganizovaného charakteru a jejich nalezení může být dílem náhody. Tento způsob zveřejňování dokumentů sice existuje, uplatňuje se, ale není zdroje prioritizovaným v rámci hnutí otevřeného přístupu.

3.4 Podpora otevřeného přístupu k informacím a jeho růst

V rozšiřování otevřeného přístupu k dokumentům hrají klíčovou roli **autoři**. Ačkoliv velké množství autorů ví o existenci volně dostupných elektronických archivů a využívá informace v nich uložené, jen málo z nich do nich aktivně přispívá, jak dokladují již realizované průzkumy [HARDY, 2005]. Velké množství autorů často využívá dokumenty, které jsou lokalizovány volně na personálních webových stránkách jejich autorů - jsou jednoduše nalézány při vyhledávání pomocí internetových vyhledávačů (Google aj.). Příznivější situace je v oblasti přírodních věd. Proces auto-archivace dokumentů v archivech či repozitářích si jen pozvolně razí cestu kupředu, nicméně vůle realizovat tento proces je u autorů průzkumy doložena. Počet autorů přispívajících do archivů se každým rokem zvětšuje. Autoři, kteří využívají auto-archivaci v otevřených archivech, oceňují jednak možnost bezplatného využívání jejich dokumentů koncovými uživateli, poskytující možnost zvyšování citovanosti jejich prací, jednak rychlost publikování. Průzkumy realizované v zahraničí také potvrzují, že velké procento autorů o možnosti auto-archivace svých prací vůbec neví. V tomto směru je pro další období důležité rozšiřování části informační gramotnosti, zejména na univerzitách.

Pro autory je důležitý vliv jejich práce. Pokud autor zjistí zlepšení **vlivu** své práce díky bezplatnému přístupu k ní, bude pravděpodobně více ochotný využívat této cesty. Mnozí odborníci uvádějí, že v některých oborech (například počítačové vědě) byla citovanost preprintů recenzovaných článků, které byly uloženy v otevřených archivech, větší než citovanost článků získaných z předplacených zdrojů. V některých průzkumech byla zjištěna vyšší citovanost otevřeně přístupných dokumentů i ve srovnání s citovaností uvedenou ve světové citační databázi WOS společnosti Thomson [HARDY, 2005].

Zvětšování počtu otevřeně přístupných zdrojů a zároveň zvětšování jejich využívání koncovými uživateli se neobejde bez **podpory řady institucí na lokální, národní i mezinárodní úrovni**. Proto ve vyspělých zemích poskytují různé výzkumné instituce a nadace nebo státní orgány mnoho finančních prostředků na rozvoj aktivit v této oblasti. Koná se velké množství konferencí a seminářů. Již dubnu roku 1992 se konalo první

Symposium o vědeckém publikování v elektronických sítích, v prosinci téhož roku pak druhé. Diskutovány byly možnosti a perspektivy v oblasti nekomerčního vědeckého publikování. Symposium bylo sponzorováno Asociací vědeckých knihoven a Asociací Amerických univerzitních nakladatelů (Association of American University Presses). Značné prostředky byly již vynaloženy ve Velké Británii (například institucí The Wellcome Trust nebo The Research Councils UK), v USA (The National Institutes of Health aj.), Finsku (Ministerstvo výchovy) aj.

Velikou podporu otevřenému přístupu k informačním zdrojům nabízejí mezinárodní organizace. Jmenujme zejména Institut otevřené společnosti (Open Society Institute), který po semináři konaném v prosinci 2001 vyhlásil 14. února 2002 **Budapešťskou iniciativu otevřeného přístupu** (BOAI, <http://www.soros.org/openaccess/>). Tato iniciativa, mimo jiné, zdůraznila potřebu auto-archivace elektronických článků do elektronických archivů a zavádění volných online časopisů. V rámci této iniciativy byly mobilizovány finanční zdroje na pomoc vzniku nových online časopisů s otevřeným přístupem a na změnu obchodních modelů u již existujících komerčních časopisů ve prospěch auto-archivace elektronických článků. Text výzvy podepsalo k současnosti již téměř **5 000 jednotlivců** a **463 organizací**. I na základě této iniciativy začal nárůst institucionálních repozitářů.

Principy a doporučení v oblasti elektronického publikování s otevřeným přístupem v biomedicínských oborech jsou kodifikovány také v **Prohlášení z Bethesdy o publikování s otevřeným přístupem** (Bethesda Statement on Open Access Publishing, <http://www.biomedcentral.com/openaccess/bethesda/>) z roku 2003. Ve stejném roce byla v říjnu přijata také **Berlínská deklarace o otevřeném přístupu ke znalostem ve vědě a humanitních oborech** (Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>), která byla podepsána hlavními německými vědeckými institucemi. Berlínská deklarace se stala impulsem k přijetí stejných principů k podpoře otevřeného přístupu v institucích dalších zemí.

Také řada **komerčních vydavatelů** se rozhodla v určité míře podporovat otevřený přístup k informačním zdrojům. Jednou z možných cest je udělit autorům svolení k auto-archivaci vlastních kopií publikovaných dokumentů (preprintů nebo postprintů). V srpnu 2002 byl odstartován projekt systému a databáze *RoMEO* (Rights Metadata for Open archiving, <http://www.sherpa.ac.uk/romeo/>), dílčího projektu *SHERPA* (Securing a Hybrid Environment for Research Preservation and Access, <http://www.sherpa.ac.uk/>), který mapuje publikační politiku komerčních vydavatelů ve vztahu k otevřenému přístupu.

Doplňková databáze poskytuje také informace o jednotlivých titulech časopisů (viz příklad záznamu na obr. 3/5), v rámci kterých je možné politiku auto-archivace realizovat (tzv. „**zelená politika**“). V rámci adaptace „zelené politiky“ dnes velký počet komerčních časopisů (vydavatelů jako jsou Elsevier, Springer aj.) dává automatické svolení k auto-archivaci - žádost autora se již nemusí předkládat oproti dřívějšímu období [SUBER, 2008b]. Někteří nakladatelé přijali hybridní politiku podpory veřejného přístupu k publikovaným dokumentům. V rámci ní požadují od autorů zaplacení poplatku za možnost zpřístupnění dokumentu ve zdroji s otevřeným přístupem. Jde o jeden z experimentálních modelů pomoci autorům zpřístupnit dokument v otevřeném zdroji. V rámci projektu SHERPA byla také již založena další databáze JULIET (<http://www.sherpa.ac.uk/juliet/>), která registruje společnosti, které finančně podporují politiku otevřeného přístupu.

23. srpna 2002 téhož roku zveřejnila také Mezinárodní federace IFLA *Internetový manifest* (<http://www.ifla.org/III/misc/internetmanif.htm>), vyzývající ke svobodnému přístupu k informacím a k odstranění bariér toku informací. Směrnice byla k Manifestu publikována 1. 12. 2006 (<http://www.ifla.org/faife/policy/iflastat/Internet-ManifestoGuidelines.pdf>). 3. 7.

2004 vydavatelství Elsevier ohlásilo jeho novou - otevřenou - politiku povolující autorům posílat konečné verze jejich plnotextových článků (uveřejněných v Elsevieru) do institucionálních repozitářů nebo na své webové stránky. 26. srpna téhož roku napsalo dvacet pět amerických laureátů Nobelovy ceny otevřený dopis Kongresu USA, ve kterém podporují plán otevřeného přístupu Národního institutu pro zdraví. 16.3. 2004 podepsalo 48 neziskových vydavatelů ve Washingtonu Principy pro otevřený přístup k vědě (<http://www.dcprinciples.org/signatories.htm>) – tento dokument k dnešnímu dni podepsalo 73 380 redakcí vědeckých časopisů a 600 000 jejích členů.

The screenshot shows the SHERPA/RoMEO website interface. At the top, the logo 'SHERPA RoMEO' is displayed in green, with the tagline '...opening access to research' and navigation links: Home | Guidance | Repositories | Projects | Links | About | Contacts. The main heading is 'Publisher copyright policies & self-archiving'. Below this, it states 'One journal found when searched for 1532-2882:'. The journal details are: Journal: **Journal- American Society for Information Science and Technology** (ISSN: 1532-2882), Publisher: **John Wiley & Sons, Ltd**. A disclaimer follows: 'These summaries are for the publishers default policies and changes or exceptions can often be negotiated by authors. All information is correct to the best of our knowledge but should not be relied upon for legal advice.' The main content area lists policies for John Wiley & Sons: Pre-print: ✓ author can archive pre-print (ie pre-refereeing); Post-print: ✓ author can archive post-print (ie final draft post-refereeing); Conditions: a bulleted list including 'On personal web site or secure external website at authors institution', 'Not allowed on institutional repository', 'JASIST authors may deposit in an institutional repository', 'Non-commercial', 'Pre-print must be accompanied with set phrase (see individual journal copyright transfer agreements)', 'Published source must be acknowledged with set phrase (see individual journal copyright transfer agreements)', 'Publisher's version/PDF cannot be used', and 'Articles in some journals can be made Open Access on payment of additional charge'; Mandated OA: a list of compliant and non-compliant options with checkmarks and crosses; Paid access: Open Choice - selected journals; Copyright: View Policy and example CTA; RoMEO: This is a RoMEO green publisher; Updated: 23-Jun-2008. Suggest an update for this record. At the bottom, there are buttons for 'Search again?', 'RoMEO colour', and 'Archiving policy', along with a browser status bar showing 'Internet' and '100%' zoom.

Obr. 3/5: část záznamu vydavatele Wiley and Sons s informací i jeho politice k otevřenému přístupu z databáze SHERPA/ROMEO [získáno 2008-12-15]

6. října 2004 společnost Google oficiálně vyhlásila projekt *Google Print* (od listopadu 2005 přejmenovaný na *Google Book Search*, <http://books.google.com>), jehož snahou je zdigitalizovat fondy největších světových knihoven a uživatelům umožnit online vyhledávání v těchto textech. 28. listopadu 2004 potom také společnost Google ohlásila vznik svého dalšího vědeckého systému *Google Scholar* (<http://scholar.google.com>). V prosinci roku 2006 se společnost Google také rozhodla kromě knih postupně digitalizovat i staré ročníky časopisů.

Složitá problematika autorsko-právních otázek volného zpřístupňování informací se stala předmětem řešení zcela nové iniciativy nazvané *Creative Commons* (<http://creativecommons.org/>), jejíž hlavní sídlo je v USA. Česká republika připravuje připojení se k této iniciativě (<http://www.creativecommons.cz/>).

3.5 *Iniciativa otevřených archivů OAI a systém ARC*

Z iniciativy představitelů nejvýznamnějších existujících digitálních archivů a digitálních knihoven světa se v říjnu 1999 konala v americkém městě Santa Fe (Nové Mexiko) významná konference, která se zabývala problematikou propojování těchto systémů a budováním nadstavbových služeb za účelem zefektivnění jejich práce zejména s ohledem na jejich koncové uživatele. Při této příležitosti byla založena virtuální komunita se jménem **Iniciativa otevřených archivů** OAI (Open Archives Initiative, <http://www.openarchives.org/>), která se věnovala a i nadále věnuje postupnému rozpracování všech otázek spojených s těmito procesy.

Jeden ze závěrů konference směřoval k přípravě závazných principů a směrnic k propojování otevřených digitálních archivů. 15. února 2000 vznikla tzv. „**Konvence ze Santa Fe pro Iniciativu otevřených archivů**„ (the Santa Fe Convention for the Open Archives Initiative, SFC). Text konvence, který je dnes historickým textem (znamenal předverzi dnes známého protokolu pro sklizení metadat OAI-PMH), je k dispozici na WWW adrese: http://www.openarchives.org/meetings/SantaFe1999/sfc_entry.htm

Konvence zahrnovala následující základní dokumenty:

1. Základní dokument Konvence ze Santa Fee
2. Definici souboru metadat OAI, tzv. OAMS
3. Soubor definovaného protokolu „Dienst“ pro komunikaci dat mezi archivy
4. Online formulář pro registraci poskytovatelů dat
5. Online formulář pro registraci poskytovatelů služeb
6. Seznam formátů producentů dat
7. Seznam přihlášených poskytovatelů dat k OAI
8. Seznam přihlášených poskytovatelů služeb

K základním digitálním archivům (jako producentů dat), které se jako první přihlásily ke spolupráci, patřily (v závorce je uveden jejich jednoznačný identifikátor, který je součástí celkového identifikátoru každého digitálního dokumentu (objektu) uloženého v celém systému propojených archivů):

1. arXiv.org e-Print archive (arXiv)
2. Cognitive Sciences Eprint Archive (CogPrints)
3. Networked Computer Science Technical Reference Library (NCSTRL)
4. National Advisory Committee for Aeronautics Technical Report Server (NACA) – NASA
5. Computer Science Teaching Center (CSTC)
6. Archivy virtuální knihovny Networked Digital Library of Theses and Dissertations (NDLTD)
7. Research Papers in Economics (RePEc)
8. Clinical Medicine and Health Research Netprints (clinmed)
9. Web Characterization Repository (WCR) a další

Původní **technický rámec** spolupráce předpokládal budování variabilních služeb pro koncové uživatele, jako jsou vědecké vyhledávací systémy (search engines), vědecké citační systémy apod. **Organizační rámec** předpokládal existenci **poskytovatelů dat** (Data providers), **poskytovatele zprostředkovatelských nadstavbových služeb** (Service providers) a mechanismus pro komunikaci dat (protokol).

Konvence ze Santa Fe byla po roce 2001 nahrazena nově vytvořeným protokolem pro sklizeň metadat OAI-PMH (Protocol for Metadata Harvesting), nejdříve ve verzi 1, a poté ve verzi 2.

Současně platná a stabilní verze je k dispozici na URL:

<http://www.openarchives.org/OAI/openarchivesprotocol.html>).

K základním principům propojování otevřených archivů patří:

1. mechanismus ukládání „jednotek“ (může to být jak digitální dokument tak metadata)
2. systémy dlouhodobého uložení a uchování jednotek
3. řízená politika uložení a ochrany digitálních jednotek
4. otevřené rozhraní ke sběru dat zprostředkovatelským systémem (poskytovatelem služby).

Poslední princip zdůrazňuje, že v současné chvíli se spolupráce orientuje na **sběr a přenos metadat** (nikoliv digitálních dokumentů samotných) od producentů dat (angl. Data Providers) směrem k poskytovatelům služeb (angl. Service Providers). Při komunikaci je využíváno **6 základních příkazů** definovaných protokolem OAI-PMH:

Příkazy informačního typu:

Identify

Zaslání požadavku na dodání základních popisných informací o archivu (repozitáři), například jeho jméno, báze URL, užívaná verze protokolu, kontakt na správce aj.

ListMetadataFormats

Zaslání požadavku na dodání seznamu všech formátů metadat, ve kterých je archiv poskytuje metadata; povinným formátem je základní specifikace Dublin Core (oai_dc)

ListSets

Zaslání požadavku na dodání seznamu kategorií tematické struktury archivu (klasifikačního schématu)

Příkazy sběrové:

ListIdentifiers

Zaslání požadavku na dodání seznamu zkrácených záznamů ve struktuře XML, záznamy zahrnují, kromě deklarací všech jmenných prostorů, v hlavičce (header) identifikátor digitálního objektu (identifier) uloženého v repozitáři, časové razítko (timestamp) a popřípadě kategorii klasifikačního schématu (set); příkaz musí být doplněn údajem o metadatovém prefixu, údajem o datu nebo údajem o kategorii klasifikačního schématu

ListRecords

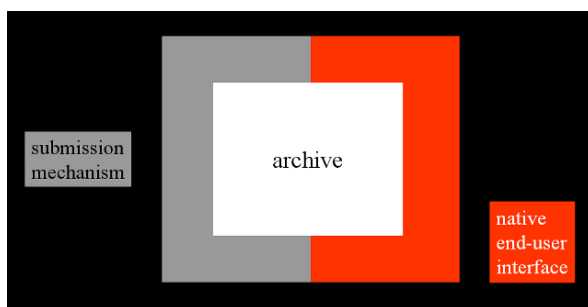
Zaslání požadavku na dodání seznamu úplných záznamů ve struktuře XML; Příkaz musí být doplněn údajem o metadatovém prefixu, údajem o datu nebo údajem o kategorii klasifikačního schématu; celý záznam zahrnuje hlavičku (Header), vlastní metadata (Metadata) a informace o původu záznamu nebo informace o možnostech využití záznamu a dokumentu (About)

GetRecord

Zaslání požadavku na dodání jednoho úplného záznamu ve struktuře XML; příkaz musí být doplněn údajem o metadatovém prefixu a údajem o identifikátoru digitálního objektu

Záznamy získávané z archivů (repozitářů) na základě protokolu OAI-PMH jsou komunikovány zásadně ve struktuře XML.

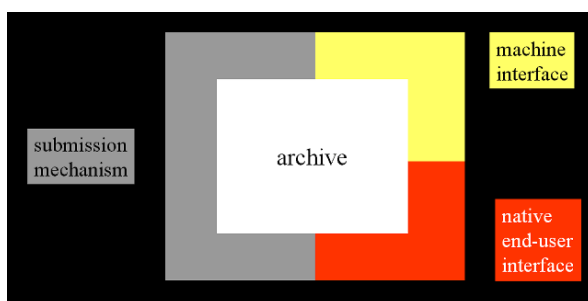
Kompletní informace k protokolu OAI-PMH pro profesionály jsou k dispozici pro zájemce na webu iniciativy OAI v základní dokumentaci tohoto protokolu [LAGOZE, 2008]. Kvalitní didaktický výklad pro začátečníky podává příručka kolektivu autorů pod vedením Leony Carpenterové [CARPENTER, 2003]. Pohodlné získání metadat z archivů, které jsou registrovány na webu OAI (viz adresy seznamu na konci této podkapitoly), je možné například ze systému „*OAI Repository Explorer*“ (<http://re.cs.uct.ac.za/>) [OAI Repository, 2006]. Tento systém byl také využit autorkou pro získání řady ukázkových metadatových záznamů, které doprovázejí části prezentující zahraniční systémy (v kapitole 4), a také pro získání záznamů dokladujících navrhované národní metadatové standardy pro komunikaci záznamů o vysokoškolských pracích z funkčních českých archivů (v kapitole 5).



Obr. 3/6: architektura typického digitálního archivu (převzato z: [Van de SOMPEL, 2000])

Na obr. 3/6-3/8 je ve schématech znázorněna základní myšlenka propojování digitálních archivů v současné době (informace jsou převzaty z článku Herberta Van de Somplela aj. – viz [SOMPEL, 2000]).

Na obr. 3/6 je znázorněn původní typický archiv elektronických tisků: zahrnuje mechanismus ukládání digitálních dokumentů včetně metadat do archivu z jedné strany (submission mechanism) a přirozené rozhraní pro koncového řešeršera a uživatele dokumentu (native end-user interface).

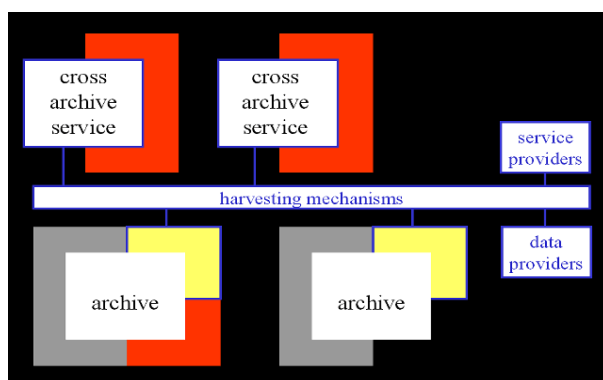


Obr. 3/7: architektura otevřeného digitálního archivu (převzato z: [Van de SOMPEL, 2000])

Na obr. 3/7 je reprezentován archiv, který zahrnuje mechanismus ukládání digitálních dokumentů včetně metadat do archivu z jedné strany a vedle přirozeného rozhraní pro

koncového uživatele i specifické rozhraní pro program (roboti, agenti), prostřednictvím něhož bude docházet ke sběru dat.

Poslední obr. 3/8 znázorňuje celkovou „scénu“ propojení otevřených digitálních archivů a jejich vazbu na zprostředkovatelské poskytovatele služby, které budou pracovat nad jednotlivými archivy



Obr. 3/8: schéma činností producentů dat a provozovatelů služeb při propojení otevřených digitálních archivů (převzato z: [Van de SOMPEL, 2000])

Prvním zaregistrovaným zprostředkovatelským systémem (registrace proběhla v říjnu 2000) v rámci OAI byla služba označená jménem **ARC** (A Cross Archive Search Service, <http://arc.cs.odu.edu/>), jejíž provoz zajišťovala Výzkumná skupina pro digitální knihovny na Univerzitě Old Dominion (Old Dominion University Digital Library Research Group, Norfolk, Virginia, <http://dlib.cs.odu.edu/>). K dnešnímu datu je v centrále OAI zaregistrováno již celkem tři desítky služeb (<http://www.openarchives.org/service/listproviders.html>). Počet zaregistrovaných archivů je v 990 (<http://www.openarchives.org/Register/BrowseSites>) ke konci března 2009.

4 Vybrané systémy digitálních archivů a digitálních knihoven s otevřeným přístupem ve vědě a výzkumu (Analýza)

V následující kapitole jsou představeny výsledky analýzy vybraných systémů elektronických archivů a digitálních knihoven z oblasti vědy a výzkumu, které jsou dnes v rámci sítě Internet uživatelům k dispozici zdarma (Open Access). Jde vesměs o systémy zpřístupňující dokumenty typu **šedé literatury**. Výběr je zaměřen na:

1. Oblast **počítačové vědy**, která má, mimo jiné, několik styčných bodů s vědou informační, a proto digitální archivy či knihovny níže představené jsou důležitým zdrojem informací i pro uživatele z oblasti informační vědy. Počítačová věda (angl. Computer Science) je navíc disciplínou, která má značné zastoupení v budovaných a provozovaných digitálních archivech a knihovnách ve světovém měřítku. Vybraní reprezentanti zastupují více typů architektury systémů, zahrnuty jsou také některé významné systémy historického charakteru. Jde o zdroje, které z několika pohledů mohou předstihovat i zdroje komerční
2. Oblast **ekonomiky**, kterou lze považovat také za obor příbuzný informační vědě; představena je rozsáhlá knihovna RePEc, která má distribuovaný charakter; zahrnuje řadu služeb potřebných pro uživatele (včetně počínajících služeb citačních) a je konkurenčním zdrojem zdrojů komerční povahy jde o výrazný konkurenční zdroj zdrojům komerčním
3. Oblast **informační a knihovní vědy**, která je zastoupena vybranými mezinárodními elektronickými archivy; jde o důležité alternativní zdroje ke zdrojům komerčním
4. Oblast **sociálních věd**, je zastoupena elektronickým archivem CogPrints
5. Elektronické knihovny typu **předmětových bran (Subject Gateways) k internetovým informačním zdrojům**, které představují neméně významný směr budování systémů pro zpracování a vyhledávání volně dostupných informací v prostoru sítě Internet z oblasti vědy a výzkumu. Jsou „lepší“ alternativou zpracování a vyhledávání informací dostupných volně v současném webu ve srovnání s vyhledávacími službami Internetu typu „search engines“ nebo univerzálních katalogů Internetu. Zvoleny jsou knihovny budované v jednom z nejvýznamnějších systémů **Intute** (dříve RDN a ROADS). Podobné produkty v současnosti zajišťují také někteří producenti významných komerčních oborových bibliografických databází (např. BIOSIS, FSTA aj.)
6. Oblast systémů centralizovaných digitálních knihoven s citačními službami nebo systémů zpracovávající citace nad elektronickými archivy
7. Oblast systémů digitálních knihoven pro zpřístupňování disertací a popřípadě i jiných vysokoškolských kvalifikačních prací

Zahrnuté systémy se dají pracovně zatřídit do následujících typů:

1. Digitální knihovny zahrnující distribuované elektronické archivy

V rámci tohoto modelu (architektury) jde o budování distribuovaných digitálních fondů a centralizovaného online přístupu k digitálním objektům pomocí společné metadatové databáze (v současné době pracují tyto knihovny vesměs všechny na bázi protokolu OAI-PMH). Příkladem mohou být:

- NCSTRL (Networked Computer Science Technical Reference Library)
(<http://www.ncstrl.org>) – v tomto případě jde již o historický systém
- NDLTD (Networked Digital Library of Theses and Dissertations)
(<http://www.ndltd.org/>)

- RePEc (Research Papers in Economics) (<http://repec.org/>)

2. Digitální knihovny typu elektronických archivů (repozitářů)

Specifický typ digitálních knihoven s výraznou komunikační a/nebo i archivační funkcí pro stanovený typ materiálů (preprinty, postprinty, popř. i vědeckovýzkumné zprávy, disertace aj.). V současnosti významně zasahují do komunikace znalostí v oblasti vybraných disciplín. Výrazným znakem je otevřený přístup (Open Access), dále auto-archivace dokumentů a zpracování metadat samotným autorem. U nejnovějších archivů je nasazováno i plnotextové zpracování uložených dokumentů. Do této skupiny patří například:

- arXiv.org e-Print archive (<http://arxiv.org/>)
- CogPrints (Cognitive Sciences Eprint Archive) (<http://cogprints.org/>)
- ECS Eprints Service (<http://eprints.ecs.soton.ac.uk>)
- E-LIS (EPrints in Library and Information Science) (<http://eprints.rclis.org>) aj.

3. Digitální knihovny typu centralizovaných kolektorů obsahů dokumentů

Model digitální knihovny založený na stahování úplných textů digitálních dokumentů z FTP archivů/repozitářů na jedno místo (server) a na jejich automatické indexaci a online vyhledávání. Specifickým příkladem je novozélandská digitální knihovna:

- NZDL (The New Zealand Digital Library, dílčí digitální fond: „Computer Science TR“) (<http://www.nzdl.org/>)

4. Předmětové brány (Subject Gateway) k internetovým zdrojům

Novodobé systémy zpracovávající pomocí ručně připravovaných kvalitních metadat informační zdroje dostupné distribuovaně v prostoru sítě Internet. Patří k nim řada systémů, například předmětové brány britského národního systému Intute (dříve RDN a ROADS) (<http://www.intute.ac.uk>)

5. Digitální knihovny typu bibliografických serverů (indexů)

Jde o online dostupné bibliografické báze dat s nebo bez přímého přístupu k úplným textům (může jít i jen o přístup k úplným textům některých dokumentů). Například:

- **HCI Bibliography** : Human-Computer Interaction Resources (<http://www.hcibib.org/>), koncem roku 2008 má databáze cca 42 000 záznamů

4.1 Elektronické archivy a digitální knihovny z oboru počítačové vědy

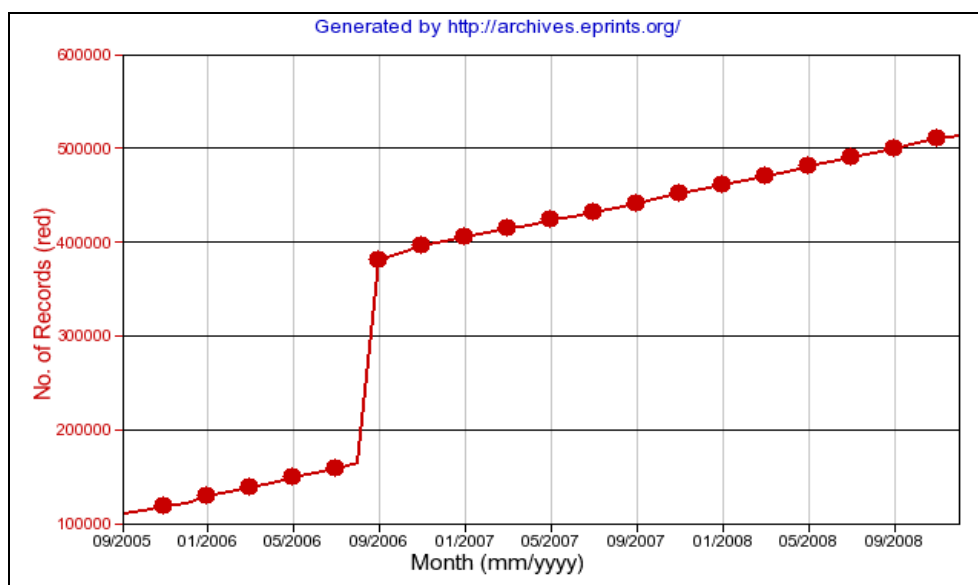
4.1.1 Archiv elektronických tisků arXiv.org

Unikátní elektronický archiv úplných textů vědecké literatury typu preprintů a postprintů, ale i vědeckovýzkumných či technických zpráv a výjimečně také disertací, **arXiv.org** (<http://arxiv.org/>) vznikl v roce 1991 v Národní laboratoři v Los Alamos (LANL, Los Alamos, New Mexico, USA). Byl dříve také znám pod zkratkou **LANL XXX** nebo jen **XXX**. Dnes je jeho sídlo na **Cornellově univerzitě** v USA, archiv vlastní a spravuje knihovna této univerzity (<http://www.library.cornell.edu/>). Dohled nad celým archivem má **Poradní rada** (Advisory Board, <http://arxiv.org/ad-board.html>), ve které je zastoupeno 15 členů, kromě jiných také **prof. Paul Ginsparg**, zakladatel archivu. Jednotlivé oborové části celého archivu mají své vlastní oborové poradní komise. ArXiv.org má k dnešnímu dni celkem **15 zrcadlových serverů** v různých částech světa (<http://arxiv.org/servers.html>).

Archiv představuje model novodobé přímé komunikace znalostí mezi vědci na bázi zcela automatizovaného informačního systému typu digitálního archivu a distribučního serveru realizovaného v prostředí sítě Internet. V době první etapy svého vývoje, kdy byla vytvářena vlastní softwarová aplikace, byl financován a podporován Ministerstvem energetiky USA (U.S. Department of Energy, Contract W-7405-Eng-36) a Národní vědeckou nadací USA (U.S. National Science Foundation, Agreement No. 9413208 - 1 Mar 1995 thru 30 Sep 2000). V současné době je jeho provoz (v nové verzi **vlastní softwarové aplikace**) zabezpečován zejména z prostředků zmíněné univerzity, částečně je dotován z prostředků Národní nadace pro vědu USA (National Science Foundation, <http://www.nsf.gov/>).

Systém zahrnuje významné obory přírodních věd: **fyziku** (hlavní a nejrozsáhlejší zastoupený vědní obor s bohatou strukturou dílčích specializací, <http://arxiv.org/>), **matematiku** (<http://arxiv.org/archive/math>), **počítačovou vědu** s návazností na počítačovou lingvistiku a neurovědu (<http://arxiv.org/archive/cs>), **nelineární vědy** (<http://arxiv.org/archive/nlin>), **kvantitativní biologii** (<http://arxiv.org/archive/q-bio>), kvantitativní finance (nově od prosince 2008, <http://arxiv.org/archive/q-fin>) a statistiku (<http://arxiv.org/archive/stat>).

Na konci roku **2008** měl archiv v repozitáři uloženo již **512 320 elektronických dokumentů** (ze všech zahrnutých oborů). Tempo růstu je od září 2006 pravidelné (viz obr. 4/1).



Obr. 4/1: graf přírůstků archivu arXiv.org za léta 2005-2008 [získáno z registru ROAR na základě dat agregátorské služby Celestial 2009-03-25]

Vstupy do systému archivu jsou realizovány prostřednictvím elektronického podávání (uploading) úplných textů včetně popisných metadat (informací o dokumentu), který v systému realizují samotní autoři dokumentů (proces auto-archivace). Před prvním uložením se autoři musejí do systému zaregistrovat (<http://arxiv.org/help/registerhelp>). Elektronické podávání dokumentů se dnes realizuje přes WWW rozhraní.

Pokud jde o formáty ukládaných textů, systém přijímá následující typy formátů (seřazeno dle priority, viz help systému <http://arxiv.org/help/submit>):

1. (La)TeX, AMS(La)TeX, PDF(La)TeX (preferovaný formát)
2. PDF
3. PostScript

4. HTML + obrázky ve formátu JPEG, PNG nebo GIF

Formáty ukládaných obrazových informací:

1. PostScript (kresby)
2. JPEG (fotografie)
3. GIF nebo PNG (nefotografický materiál)
4. PDF

Autoři dokumentů při podávání dokumentů vyplňují stručné metadatové záznamy o svých dokumentech. Záznamy jsou tvořeny podle stanovených zásad v souladu s internetovým standardem RFC 1807 „*A Format for Bibliographic Records*“ (text normy viz např. na adrese: <<http://www.ietf.org/rfc/rfc1807.txt>>). Stručné ale výstižné pokyny k vyplňování metadat mají autoři k dispozici v rámci helpu systému (<http://arxiv.org/help/prep>). Metadatový záznam v aplikaci archivu zahrnuje zejména, pokud jsou uloženy, následující prvky:

- Název dokumentu (Title)
- Autoři (Authors)
- Abstrakt (Abstract)
- Poznámky (Comments)
- Číslo zprávy (jde-li o výzkum) (Report-no)
- Předmětovou skupinu (z nabízené hrubé kategorizace archivu arXiv.org) (Subj-class)
- Informace o časopise, který publikoval článek týkající se zprávy (Journal-ref)
- DOI (identifikátor, existuje-li – viz příklad 4/2)
- Třídění MSC (Mathematics Subject Classification) - pouze pro matematiku (MCS-class)
- Třídění AMC (Association for Computing Machinery) (ACM-class)

Příklad záznamu preprintu článku (určeného pro časopis) z oboru počítačové vědy s identifikátorem „oai:arXiv.org:cs/0612140“ v metadatovém formátu „arXiv“ (ve struktuře XML), vyexportovaného z repozitáře pomocí protokolu OAI-PMH, je na obr. 4/2.

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2009-03-26T16:19:20Z</responseDate>
  <request verb="GetRecord" identifier="oai:arXiv.org:cs/0612140"
    metadataPrefix="arXiv">http://export.arxiv.org/oai2</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:arXiv.org:cs/0612140</identifier>
        <timestamp>2009-03-23</timestamp>
        <setSpec>cs</setSpec>
      </header>
      <metadata>
        <arXiv xmlns="http://arxiv.org/OAI/arXiv/"
          xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
          xsi:schemaLocation="http://arxiv.org/OAI/arXiv/
            http://arxiv.org/OAI/arXiv.xsd">
          <id>cs/0612140</id>
          <created>2006-12-28</created>
          <updated>2007-03-28</updated>
          <authors>
```

```

=<author>
  <keyname>Barbosa</keyname>
  <forenames>Valmir C.</forenames>
</author>
=<author>
  <keyname>Ferreira</keyname>
  <forenames>Fernando M. L.</forenames>
</author>
=<author>
  <keyname>Kling</keyname>
  <forenames>Daniel V.</forenames>
</author>
=<author>
  <keyname>Lopes</keyname>
  <forenames>Eduardo</forenames>
</author>
=<author>
  <keyname>Protti</keyname>
  <forenames>Fabio</forenames>
</author>
=<author>
  <keyname>Schmitz</keyname>
  <forenames>Eber A.</forenames>
</author>
</authors>
<title>On simulating nondeterministic stochastic activity
networks</title>
<categories>cs.DM</categories>
<journal-ref>European Journal of Operational Research 198
(2009), 266-274</journal-ref>
<doi>10.1016/j.ejor.2008.06.010</doi>
<abstract>In this work we deal with a mechanism for process
simulation called a NonDeterministic Stochastic Activity
Network (NDSAN). An NDSAN consists basically of a set of
activities along with precedence relations involving these
activities, which determine their order of execution. Activity
durations are stochastic, given by continuous, nonnegative
random variables. The nondeterministic behavior of an
NDSAN is based on two additional possibilities: (i) by
associating choice probabilities with groups of activities,
some branches of execution may not be taken; (ii) by
allowing iterated executions of groups of activities according
to predetermined probabilities, the number of times an
activity must be executed is not determined a priori. These
properties lead to a rich variety of activity networks, capable
of modeling many real situations in process engineering,
project design, and troubleshooting. We describe a recursive
simulation algorithm for NDSANs, whose repeated execution
produces a close approximation to the probability distribution
of the completion time of the entire network. We also report
on real-world case studies.</abstract>
</arXiv>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

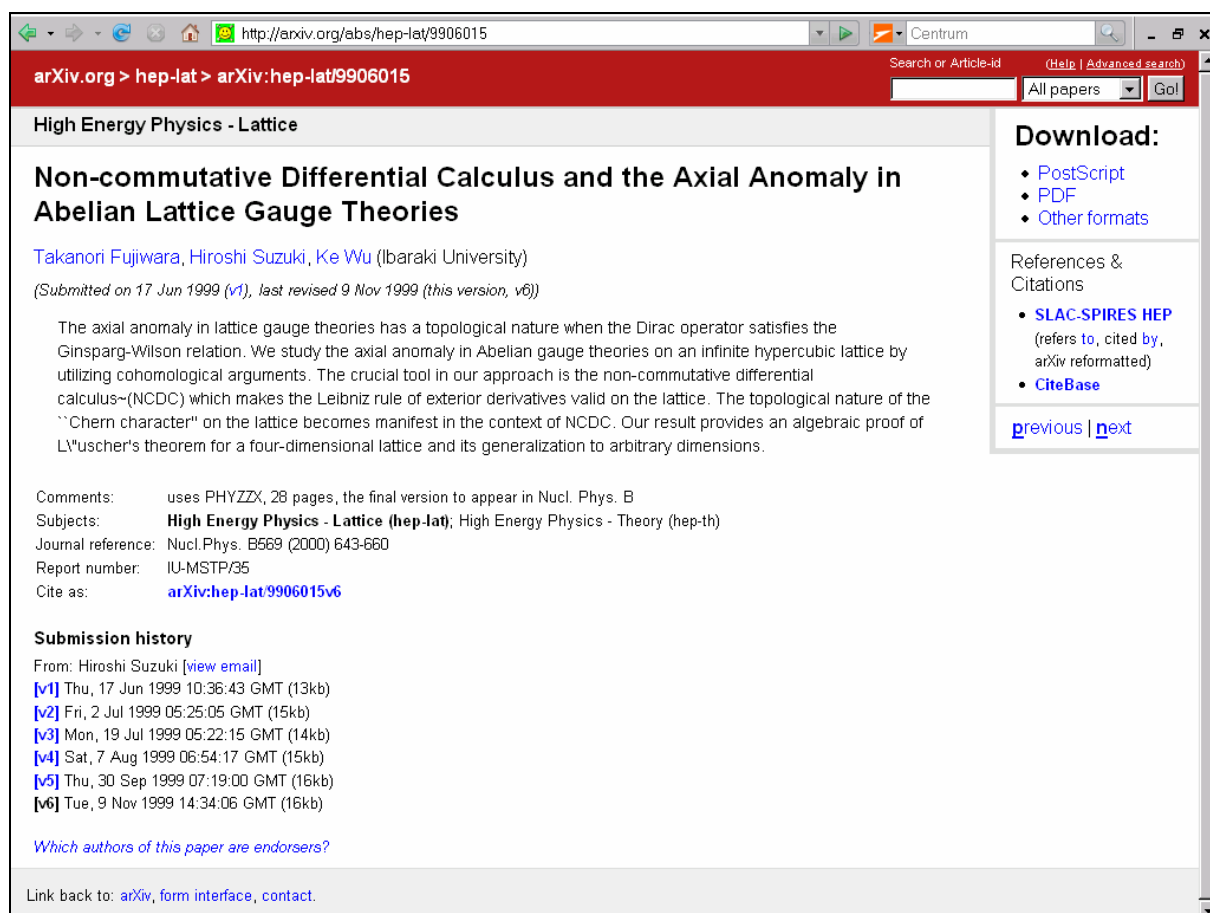
```

Obr. 4/2: metadatový záznam z archivu arXiv.org ve struktuře XML exportovaného pomocí protokolu OAI-PMH pomocí příkazu GetRecord (ve formátu arXiv) [převzato 2009-03-25]

Po zpracování metadat v systému je každému digitálnímu dokumentu (bez ohledu na jeho verze) přidělen jedinečný identifikátor. Od 1. 4. 2007 jsou přidělovány identifikátory v následujícím modelu, příklad:

arXiv:0706.0001v1

Návěští „arXiv“ je odděleno dvojtečkou od dvouznakového identifikátoru pro rok uložení dokumentu (poslední dvojčíslí roku) a dále od měsíce uložení (dvojčíslí). Po syntaktické teče následuje sekvenční číslo (4 číslice) pro daný digitální dokument (v jednom měsíci může být teoreticky uloženo 9999 dokumentů, předpokládá se případné zvětšení čísla na 5 znaků). Konec identifikátor může zahrnovat informaci o verzi daného digitálního dokumentu. Verzovitost jednoho dokumentu lze dokladovat na příkladu záznamu na obr. 4/3 – údaje o verzích jsou prezentovány na konci záznamu za návěštím „Submission history“. Příklady identifikátorů staršího modelu (zahrnovaly písmenné označení částí archivů) lze vidět také v záznamu na obr. 4/3 (arXiv:hep-lat/9906015v6).



arXiv.org > hep-lat > arXiv:hep-lat/9906015

High Energy Physics - Lattice

Non-commutative Differential Calculus and the Axial Anomaly in Abelian Lattice Gauge Theories

Takanori Fujiwara, Hiroshi Suzuki, Ke Wu (Ibaraki University)

(Submitted on 17 Jun 1999 (v1), last revised 9 Nov 1999 (this version, v6))

The axial anomaly in lattice gauge theories has a topological nature when the Dirac operator satisfies the Ginsparg-Wilson relation. We study the axial anomaly in Abelian gauge theories on an infinite hypercubic lattice by utilizing cohomological arguments. The crucial tool in our approach is the non-commutative differential calculus-(NCDC) which makes the Leibniz rule of exterior derivatives valid on the lattice. The topological nature of the "Chern character" on the lattice becomes manifest in the context of NCDC. Our result provides an algebraic proof of Lüscher's theorem for a four-dimensional lattice and its generalization to arbitrary dimensions.

Comments: uses PHYZZX, 28 pages, the final version to appear in Nucl. Phys. B

Subjects: **High Energy Physics - Lattice (hep-lat)**; High Energy Physics - Theory (hep-th)

Journal reference: Nucl.Phys. B569 (2000) 643-660

Report number: IU-MSTP/35

Cite as: [arXiv:hep-lat/9906015v6](https://arxiv.org/abs/hep-lat/9906015v6)

Submission history

From: Hiroshi Suzuki [view email]

- [v1] Thu, 17 Jun 1999 10:36:43 GMT (13kb)
- [v2] Fri, 2 Jul 1999 05:25:05 GMT (15kb)
- [v3] Mon, 19 Jul 1999 05:22:15 GMT (14kb)
- [v4] Sat, 7 Aug 1999 06:54:17 GMT (15kb)
- [v5] Thu, 30 Sep 1999 07:19:00 GMT (16kb)
- [v6] Tue, 9 Nov 1999 14:34:06 GMT (16kb)

[Which authors of this paper are endorsers?](#)

Link back to: [arXiv](#), [form interface](#), [contact](#).

Download:

- [PostScript](#)
- [PDF](#)
- [Other formats](#)

References & Citations

- [SLAC-SPIRES HEP](#) (refers to, cited by, arXiv reformatted)
- [CiteBase](#)

[previous](#) | [next](#)

Obr. 4/3: příklad záznamu 6. verze dokumentu s identifikátorem (starší typ) arXiv:hep-lat/9906015 (archiv fyziky elementárních částic) v úplném výstupním formátu (zahrnuje odkazy na předchozí verze a také odkazy na citované a citující autory, získáno 2008)

Vyhledávání informací

Archiv arXiv.org má podle denně získávaných statistik vysokou **návštěvnost** ze strany uživatelů z celého světa. Počet spojení denně v posledním období se pohybuje kolem 150 000-200 000.

Základní vyhledávání v digitálním archivu je realizováno přes index, který je připravován na základě uložených metadat (včetně abstraktu). S ohledem na menší počet údajů v záznamech metadat není nabídka selekčních údajů pro vyhledávání rozsáhlá. Vyhledávání v úplných textech je realizováno již v systému také (viz přístup z rozhraní pro pokročilé vyhledávání (<http://arxiv.org/find>)).

V rámci systému lze vyhledávat dvojím základním způsobem:

- prohlížením seznamů záznamů dokumentů v dílčích tematicky rozdělených archivech, popř. prohlížení podle tematických skupin a dále podle data uložení
- přímým vyhledáváním přes formulaci dotazu; selekčními údaji jsou: identifikátor dokumentu, číslo výzkumné zprávy, předmětová skupina, název dokumentu, autor, poznámky, název časopisu, třídník, klíčová slova z abstraktu. Od září 1999 systém **podporuje** booleovské vyhledávání v jednotlivých polích záznamů, podporováno je také vyhledávání frází, automatické rozšíření selekčního údaje apod.

Úplné záznamy dokumentů v prezentačních formátech zahrnují v tuto chvíli již řadu důležitých **hypertextových odkazů**, pomocí nichž lze navigovat mezi záznamy nebo k jejich množinám. Navigovat lze také přes hypertextový odkaz k záznamům jiných verzí téhož dokumentu (mají stejný identifikátor) nebo k seznamu záznamů dokumentů jednoho autora. Ze záznamu se lze po volbě výstupního formátu dostat k úplnému textu. Podstatnou vlastností jsou funkční hypertextové odkazy ze záznamů citací v úplném textu na záznamy citovaných dokumentů, pokud jsou součástí archivu (viz obr. 4/3 „References & Citations“). U některých záznamů jsou již také v abstraktu vidět hypertextové odkazy na jiné záznamy dokumentů uložených v archivu.

Důležitou službou je služba zobrazování jednak **seznamu citovaných dokumentů** v daném dokumentu, jednak **seznamu citací** na daný dokument, kterou zajišťuje externě známá databáze **SPIRES-HEP** (<http://www.slac.stanford.edu/spires/index.shtml>), provozovaná Stanfordskou univerzitou (USA). Seznamy se vyvolávají z odkazů v záznamech (**refers to a cited by** - viz taktéž na obr.4/3). Jde však o vlastnosti, které jsou předmětem dalšího zdokonalování v nových projektech. Další citační služba je zajišťována již výše zmíněným systémem **Citebase Search** (<http://www.citebase.org/>) [BRATKOVÁ, 2006, s. 13].

Úplné texty dokumentů včetně abstraktů lze získat z archivu zdarma a zpravidla ve více formátech. Nabídka výstupních formátů je ovlivněna vstupním formátem, který dodal autor dokumentu. Běžné je zobrazování (s možností stažení) úplného textu ve formátu zdrojovém (varianty TeX), formátu PostScript nebo jiných formátech (DVI, PDF aj.).

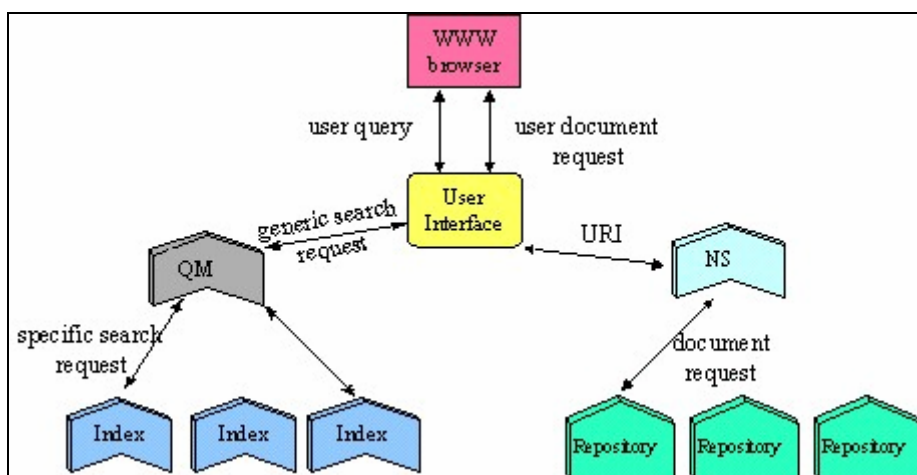
Od roku 1998 byl po dohodě tří významných amerických partnerů (vydavatelské organizace ACM, původního elektronického archivu XXX v Los Alamos a dřívější distribuované digitální knihovny NCSTRL) vybudován na bázi osvědčeného programového systému archivu XXX (dnes arXiv.org) **kooperativní archiv z oblasti počítačové vědy**, který získal označení **CoRR (Computing Research Repository)** - viz dále v této kapitole).

4.1.2 NCSTRL (Networked Computer Science Technical Reference Library)

Síťová technická referenční knihovna pro počítačovou vědu **NCSTRL** (vyslovuje se „ancestral“, původní URL <http://www.ncstrl.org/>) byla ve své době jednou z nejvýznamnějších a největších distribuovaných digitálních knihoven na světě. Její fond byl tvořen elektronickými výzkumnými a technickými zprávami, konferenčními referáty, ale také články z elektronických časopisů dostupných na WWW z oboru **počítačové vědy**. Digitální fond zahrnoval textové dokumenty pocházející z téměř 100 kooperujících institucí z celého světa (Severní Ameriky, Asie, Evropy aj.). Šlo většinou o univerzity realizující doktorská studia v oblasti počítačové vědy nebo v příbuzných inženýrských oborech. Systém byl připravován, realizován a rozvíjen od roku 1994 na Cornellově univerzitě (Cornell University) týmem odborníků sdružených v rámci skupiny „Cornell Digital Library Research Group“ (CDLRG), Ithaca (N.Y), USA. Finančně byl projekt v prvním období podporován agenturou ARPA (Advanced Research Projects Agency - Grant No. MDA972-92-J-1029),

dále CNRI (Corporation for National Research Initiatives) a NSF (National Science Foundation - Grant No. NSF-CDA-9308259.).

Pokud šlo o model a technologie, byl systém NCSTRL budován jako globální distribuovaná digitální knihovna. V prvním období do roku 2001 šlo o **architekturu federace** distribuovaných FTP archivů s příslušnými lokálními indexy metadat. Vyhledávání informací bylo zajišťováno virtuálně centrálním systémem. Interakce mezi jednotlivými subsystémy a přístupem k dokumentu byla zajišťována přes bývalý **protokol Dienst** (Dienst Protocol). Slovo **Dienst** v germánských jazycích sice označuje „službu“, avšak z hlediska systémového a technologického mělo širší význam: Vedle vlastního komunikačního protokolu označoval i programový systém pro jeho implementaci a konečně označoval i celou koncepci a architekturu digitální knihovny.



Obr. 4/4: schéma původní interakce služby Dienst v digitální knihovně NCSTRL [LAGOZE, 2000]

Schéma na obr. 4/4 představuje původní **architekturu knihovny NCSTRL**. Dotaz uživatele pracujícího s webovým prohlížečem (kdekoliv v Internetu) byl zapisován do formuláře v rámci uživatelského rozhraní (User Interface) lokalizovaného na centrálním serveru celé digitální knihovny (Cornell University). Formulace dotazu byla v této fázi předávána přes standardní protokol HTTP. Zprostředkovatel dotazu (QM Query Mediator) zabezpečoval překlad a distribuci dotazu všem účastnickým systémům (jejich indexům). Výsledky vyhledávání v indexech se vracely zpět do QM ke zpracování (uživatel je dostával ve formě stručných bibliografických záznamů - níže viz na obr. 4/5 příklad jednoho záznamu z rešerše). Pro získání úplného textu dokumentu byl nezbytný identifikátor digitálního objektu (Handle, resp. URI, ve výstupních záznamech je označen jako „Document ID“), na základě kterého byl prostřednictvím identifikačního systému (NS Name Server) v příslušných distribuovaných repozitářích nalezen, kopírován a v konečné fázi uživateli dodáván úplný text. Uživatel si mohl stáhnout dokument v jednom z nabízených formátů.

Federativní model byl od roku 2001 postupně nahrazován novým modelem založeným na **protokolu OAI-PMH**, tj. metadata z některých repozitářů byla harvesterem sklízena do centrální databáze (byla provozována v systému ORACLE), nad kterou bylo také centrální vyhledávací rozhraní (bylo lokalizováno na serveru Univerzity Old Dominion, která také v tehdy provozovala **službu ARC**, <http://arc.cs.odu.edu>) - informace o této službě jsou uvedeny v další části textu. Z některých archivů, které do té doby nebyly připraveny na sklizení podle protokolu OAI-PMH, se data musela získávat náhradní cestou. Lokální pracoviště se vybavovala novými programy pro budování elektronického archivu (například

programem Eprints, <http://www.eprints.org>), které zabezpečují bezproblémové sklizení metadat pro centrální databázi.

Celý digitální fond zahrnoval úplné texty dokumentů (výzkumných a technických zpráv, konferenčních referátů, ale i článků).

Networked Computer Science Technical Reference Library

Simple Search Advanced Search Browse Register Submit to CoRR About NCSTRL OAI Help

SEARCH RESULTS GROUPED BY ARCHIVE

This is page 1 of total 6 pages, hits (1--10) of total 53 hits.

Results Pages: 1 2 3 4 5 6

ARCHIVE : CNRI

Title [The NSF National Science, Mathematics, Engineering, and Technology Education Digital Library \(NSDL\) Program: A Progress Report](#)

Authors Zia, Lee L.

Archive CNRI

Discovery Date 2000-10-16

Document ID oai:ncstrl:cnri_dlib:cnri_dlib/october2000-zia

Title [Working Towards an Understanding of Digital Library Use: A Report on the User Research Efforts of the NSF/ARPA/NASA DLI Projects](#)

Authors Bishop, Ann Peterson

Archive CNRI

Discovery Date 1995-10-15

Document ID oai:ncstrl:cnri_dlib:cnri_dlib/october95-bishop

Title [DigitalLSI: A Digital Library System Based on Hyper-G](#)

Authors Tochtermann, Klaus | Alders, Thomas

Archive CNRI

Discovery Date 1996-10-15

Document ID oai:ncstrl:cnri_dlib:cnri_dlib/october96-tochtermann

Note: IE 6.0 users, please enable third party cookies

This is a collaborative project involving [NASA Langley](#), [Old Dominion University](#), [University of Virginia](#) and [Virginia Tech](#).

Powered by **arc** [Old Dominion University](#)

Obr. 4/5: seznam zkrácených záznamů ze systému knihovny NCSTRL ve výstupním formátu [získáno 2006-01-02]

Knihovna NCSTRL přijímala následující typy **formátů primárních textových dokumentů** (seřazeno dle priority):

1. PostScript (základní požadovaný formát pro uložení NCSTRL)
2. HTML
3. PDF

K preferovaným **formátům obrazových informací** patřily formáty:

1. TIFF
2. GIF

Základní instrukcí k tvorbě **metadat** pro účastnické knihovny byl také internetový standard RFC 1807 „A Format for Bibliographic Records“ (text normy viz např. na adrese: <http://www.ietf.org/rfc/rfc1807.txt>). Množství údajů se ovšem lišilo v jednotlivých institucích. Jistým problémem, jak vyplývalo z publikovaných analýz, byla nejednotná forma zápisu údajů (např. jmen autorů). K základním údajům patřily:

- Identifikátor dokumentu
- Název dokumentu
- Autoři
- Abstrakt

Vyhledávání informací

S ohledem na menší počet údajů v záznamech metadat, nebyla nabídka selekčních údajů pro vyhledávání i v tomto případě bohatá. Vyhledávání v úplných textech nebylo realizováno.

V rámci systému bylo možné jednak prohlížet záznamy v rámci jednotlivých distribuovaných archivů (Browse), jednak vyhledávat pomocí přímé formulace dotazu; a to v rámci **rozhraní jednoduchého** (Simple Search) nebo **pokročilého** (Advanced Search). Prohledávat bylo možné všechna pole záznamů (autor, název abstrakt). V rámci polí i mezi polí bylo možné užívat operátory AND a OR a také uvozovek pro hledání frází. Po zadání fráze bez uvozovek systém hledal automaticky pomocí operátoru AND. Pro složitější dotazy bylo možné užívat kulatých závorek. Zadaná slova byla při vyhledávání automaticky rozšiřována (lemmatizace). Při formulaci nebylo možné zapisovat jednotlivá písmena, gramatické členy apod. („a“, „s“, „the“ ...). Výsledný příkladový zkrácený záznam je na obr. č. 4/5.

Úplné texty dokumentů včetně abstrakt bylo možné získat zpravidla ve více formátech. Nabídka výstupních formátů byla ovlivněna vstupním formátem. Běžné bylo zobrazování (s možností stažení) úplného textu ve formátu PostScript nebo PDF. Systém poskytoval také pro zájemce službu SDI (přes e-mail).

Digitální referenční knihovna NCSTRL z důvodu zastaralé koncepce systému ukončila svoji činnost koncem roku 2007 (viz zpráva na WWW: <http://sunsite.berkeley.edu/NCSTRL/>). Bývalé URL (<http://www.ncstrl.org/>) bylo určitý čas nedostupné v současné chvíli je hlavní stránka rozhraní ještě dostupná, vyhledávání již ale není funkční. Potěšitelná je ale zpráva, že všechny elektronické dokumenty knihovny NCSTRL byly transferovány do moderního archivu elektronických tisků (technických zpráv) „**EECS Technical Reports**“ (<http://techreports.lib.berkeley.edu/>), který spravuje Fakulta elektrického inženýrství a počítačové vědy na Kalifornské univerzitě v Berkeley (Department of Electrical Engineering and Computer Science, University of California, Berkeley, USA, <http://www.eecs.berkeley.edu/>).

4.1.3 CoRR (Computing Research Repository)

V roce 1998 po jednání a dohodě zástupců vydavatelské organizace **ACM** (Association for Computing Machinery), původního archivu elektronického tisku **XXX** v Los Alamos (nyní arXiv.org) a bývalé digitální knihovny **NCSTRL** na Cornellově univerzitě vznikl a do provozu byl uveden nový kooperativní systém elektronického archivu pro obor **počítačové vědy**, jehož cílem bylo co nejrychlejší archivování výzkumných zpráv v centrálním repozitáři a jejich okamžité zpřístupňování uživatelům. Nový archiv dostal jméno **CoRR (Computing Research Repository, tj. Repozitář výzkumu v počítačové vědě)**. Měl představovat sjednocení tehdy početných repozitářů a FTP archivů. Starší hlavní webová stránka původního systému je stále k dispozici (<http://www.acm.org/corr/>). Aktuální systém (je dnes součástí archivu arXiv.org) se nalézá na jiné adrese (<http://arxiv.org/corr/home>).

Na základě implementace známého protokolu Dienst **CoRR** původně kombinoval otevřenou architekturu digitální knihovny NCSTRL se spolehlivými a osvědčenými technologickými praktikami přístupu (vstupy, ukládání i výstupy) do elektronického archivu **arXiv.org**. Tato

architektura umožňovala integraci s jinými archivy a pokládala tak základ pro budoucí široce založenou vědeckou digitální knihovnu. V současné době je systém CoRR provozován již na serveru archivu arXiv.org.

CoRR provozuje repozitář, do kterého všichni členové komunity mohou nezávisle předkládat konferenční referáty a výzkumné zprávy, ale dnes také další typy dokumentů. Zároveň CoRR nabízí uživatelské přívětivé rozhraní, pomocí něhož mohou v archivu vyhledávat i jiní uživatelé. Ti se mohou z druhé strany stát po registraci novými autory.

Počet uložených dokumentů dosáhl od svého založení (1998-) počtu cca **80 000 uložených dokumentů**, z toho 900 dokumentů bylo převedeno z původního archivu LANL XXX - tematické oblasti počítačové vědy (včetně produkce za léta 1993-1997). Nové přírůstky přibývají v objemech cca 25 000 dokumentů ročně. Denně je realizováno cca 70 000 transakcí se systémem.

Problémy řešené v období budování CoRR:

1. Třídění dokumentů:

Příslušná společná komise musela v té době rozřešit kompromisem otázku věcného pořádání dokumentů. ACM požadovala své standardní třídění (ACM Classification), které zahrnuje 100 předmětových kategorií uspořádaných hierarchicky ve třech úrovních, představitelé archivu arXiv.org navrhli vlastní třídění s 33 základními kategoriemi (pro každou je stanoven jeden moderátor-editor, který kontroluje po autorech správnost zařazení jejich dokumentu). Autoři byli povinováni zařadit svůj dokument do obou třídění.

2. Copyright:

Příslušná komise rozhodla, že autoři, kteří ukládali dokumenty do repozitáře, si ponechají copyright, resp. jinou formu autorských práv. Tato otázka se však musela řešit i nadále, protože řada autorů navíc publikovala svoje výzkumy i v komerčních časopisech či sbornících, takže mohla narážet na podmínku předávání copyrightu. Většina vydavatelů je již dnes tolerantních, realizují tzv. „zelenou politiku“.

3. Doba uložení dokumentu v archivu:

CoRR měl mít výraznou archivní funkci, tj. předpokládalo se, že dokument (výzkumná zpráva, konferenční referát, preprint atd.) zůstane trvale v archivu. Autoři nemohli měnit již jednou uložené dokumenty, resp. nahrazovat je aktuálními verzemi, mohli však připojovat verze nové (dokumenty jsou označovány datem uložení, takže nemohlo dojít k záměně). Nejnovější verze je uživateli nabízena běžně jako první, v systému však jsou zabudovány křížové odkazy, které ho ke starším verzím dovedou.

4. Formát digitálního objektu:

Otázka formátu, ve kterém autoři dodávají své dokumenty, byla a je jednou z nejvíce diskutovaných. Vzhledem k tomu, že v elektronickém archivu arXiv.org je preferován formát TeX (jeho různé varianty), bylo rozhodnuto prozatím zůstat u preference tohoto formátu i v rámci CoRR. Autoři nemohou odevzdávat dokument např. v oblíbeném PDF formátu, pokud existuje původní forma ve zdrojovém formátu TeX. Systém na základě zdrojového formátu poskytuje uživatelům různé výstupní formáty na bázi automatické konverze: různé platformy formátu PostScript, PDF aj.

5. Organizační a finanční zabezpečení

Otevřeným problémem byla otázka **organizace** celého systému (palčivý byl vztah CoRR k bývalému systému NCSTRL). Autoři mnoha institucí, které byly členy bývalé knihovny

NCSTRL, dříve ukládali své dokumenty v repozitářích, resp. v FTP archivech svých institucí. Později je začali ukládat do CoRR. Otázka zněla: co udělají další archivy? Dojde k reorganizaci systému NCSTRL? Na řešení v dalších letech působily i další faktory ovlivňující současnou komunikaci vědeckých informací ve společnosti. CoRR je dnes již trvalou součástí systému arXiv.org. Palčivou otázkou bylo také **finanční zabezpečení**. Elektronické archivy arXiv.org byly sponzorovány ze strany různých nadačních organizací USA podporujících výzkum. Nový archiv CoRR sponzoruje v tuto chvíli také ACM.

Základní principy a postupy (registrace autorů, ukládání dokumentů, tvorba metadat, vyhledávání informací, výstupy úplných textů atd.) v elektronickém archivu CoRR jsou stejné jako v archivu arXiv.org, proto nebudou v tomto místě opakovány.

4.1.4 Institucionální repozitář elektronických tisků ECS

4.1.4.1 Úvodní charakteristika repozitáře

V titulku jmenovaný archiv je zástupcem kategorie „institucionální repozitáře“. Zkratka **ECS** patří Škole elektrotechniky a počítačové vědy (School of Electronics & Computer Science, <http://www.ecs.soton.ac.uk/>), která náleží do struktury Southamptonské univerzity (University of Southampton, <http://www.soton.ac.uk/>). Jde o renomované, úspěšné pracoviště na území Velké Británie v oblasti výzkumu a výuky výše zmíněných oborů.

ESC spravuje vlastní (stejnojmenný) institucionální repozitář (IR) elektronických tisků (<http://eprints.ecs.soton.ac.uk/>), který je významný co do obsahu pro uživatele z celého světa (texty pokrývají zejména realizované výzkumy a vývoj a vynikají vysokou kvalitou). Jde o jeden z největších institucionálních repozitářů na světě, v prosinci 2008 zahrnoval úplné texty s metadatovými popisy cca **13 500** dokumentů. Škola je zapojena aktivně v oblasti elektronických archivů v globálním rámci. Byla jednou z prvních institucí, která již v roce 2001 nařídila svým členům archivovat dokumenty vzniklé v rámci univerzity.

I to byl důvod, proč byl brzy na této škole vytvořen jeden z nejvyužívanějších softwarů pro potřeby elektronických archivů - software **EPprints**. (<http://www.eprints.org/>). Autorem verze 1 byl R. Tansley, verzi 2 a verzi 3 již zajišťuje Ch. Gutteridge s dalšími spolupracovníky. Rozvoj programu je finančně podporován britskou Společnou komisí pro informační systémy JISC (<http://www.jisc.ac.uk/>) a také americkou Národní vědeckou nadací NSF (<http://www.nsf.gov/>). Software je k dispozici zdarma a je určený pro nekomerční využití. Program ve své nejnovější verzi EPprints 3 nabízí pohodlné a přívětivé uživatelské rozhraní a sadu standardních služeb (vyhledávání informací v různých režimech, propojení jednotlivých částí systému, registraci i jednoznačnou identifikaci uživatelů aj.). Podrobnější popis tohoto programu je uveden v části 4.3.1. Program je i nadále rozvíjen a inovován.

4.1.4.2 Obsahové vymezení repozitáře ECS

Repozitář ECS se specializuje na obor elektroniky, elektroinženýrství, počítačové věda včetně počítačových a informačních technologií. Bližší obsah prozrazuje rejstřík repozitáře „Výzkumné skupiny“ (Browse by Research Groups), který nabízí **22 skupin** (v kulaté závorce je počet dokumentů náležejících do dané skupiny ke konci roku 2008):

- BIO@ECS Research Group (65)
- Communications Research Group (1554)
- Dependable Systems and Software Engineering Research Group (1119)
- Electrical Power Engineering (1108)
- Electronic Systems Design Group (909)

- Electronics Construction, Maintenance and Logistics (6)
- Grid and Pervasive Computing Research Group (8)
- High Performance Computing Research Group (396)
- Information – Signals, Images, Systems (1849)
- Infrared Science and Technology (188)
- Intelligence, Agents, Multimedia (2746)
- IT Innovation Centre (186)
- Learning Societies Lab (360)
- Mechanical Workshop (1)
- Microelectronics Industrial Unit (2)
- Nanoscale Systems Integration Group (1693)
- No Specific Group (212)
- Optoelectronics Research Centre (1231)
- Pervasive System Centre (27)
- Science and Engineering of Natural Systems (350)
- Systems and Networks Group (3)
- Systems Group (10)
- University of Southampton Institute of Transducer Technology (2)

4.1.4.3 Typologie dokumentů

Institucionální archiv ECS přijímá následující hlavní typy dokumentů:

- Články (Article)
- Materiály z konferencí či workshopů (Conference or Workshop Item)
- Knihy (Book)
- Části knih (Book Section)
- Technické zprávy (Technical Report)
- Disertační práce (Thesis)
- Jiné dokumenty (Other)

4.1.4.4 Vyhledávání informací

Institucionální repozitář ECS nabízí ve verzi EPrints 3 více možností vyhledávání informací. Pro nezaškolené uživatele je k dispozici jednoduché vyhledávání pomocí **prohlížení rejstříků** systému: podle již výše uvedené předmětové či oborové kategorie (Browse by Research Groups) a podle roku zveřejnění dokumentu (Browse by Year) aj. Po kliknutí na vybranou položku rejstříku zobrazí systém seznam položek podle typů dokumentů a dále podle roku jeho zveřejnění. V rejstříku roků zveřejnění dokumentů se po kliknutí vybraný rok zobrazí seznam krátkých záznamů, které jsou rozděleny podle typu dokumentu.

Proces prohlížení rejstříku končí zobrazením úplného záznamu dokumentu (zahrnuje název dokumentu, podrobný abstrakt, hyperlink na úplný text, a jiné údaje).

Dotazové vyhledávání zahrnuje režim **jednoduchý** (Search), který je dostupný z mnoha stránek systému, a režim **pokročilý** (Advanced Search).

Rozhraní pro **jednoduché vyhledávání** zahrnuje možnost nastavit uspořádání nalezených výsledků (*Order the results*) podle roku, jména autora nebo názvu dokumentu.

Pokročilé vyhledávání nabízí velkou řadu selekčních údajů (předmětová kategorie, typ dokument, stav publikování dokumentu aj.). Uživatel zadává dotaz pomocí pohodlného předdefinovaného formuláře.

Výsledné vyhledané záznamy se zobrazují nejdříve ve zkrácené verzi, která umožňuje přes hyperlinkovaný název dokumentu získat záznam úplný včetně abstraktu. Odkaz na úplný text má novou vlastnost – ikona PDF při ukázání myši nabízí náhled na titulní stránku dokumentu. Jména autorů obsahují hyperodkaz na záznam autora a kompletní seznam jeho dokumentů v archivu. Další hyperodkaz z úplného záznamu je veden z předmětové či tematické kategorie. Součástí úplného záznamu je také denní a měsíční statistika stahování (Download Statistics). Většina záznamů je doplněna seznamem použité literatury, kterou zpracovává návazná citační služba systému Citebase Search.

Dotazový vyhledávací režim nabízí novinku 3. verze softwaru EPrints - záznamy lze exportovat do celé řady metadatových formátů, například:

- BibTex
- Dublin Core
- DIDL
- EndNote
- HTML Citation
- METS
- MODS
- Simple Metadata
- EP3 XML

Obr. 4/6 je záznam publikace (typu výzkumné zprávy), exportovaný pomocí automatické funkce rešeršního rozhraní EPrints do metadatového formátu MODS.

```
-----
<?xml version="1.0" encoding="UTF-8" standalone="yes">
<mods:mods xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:mods="http://www.loc.gov/mods/v3" version="3.0"
  xsi:schemaLocation="http://www.loc.gov/mods/v3
    http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
  <mods:titleInfo>
    <mods:title>Open Access Citation Information</mods:title>
  </mods:titleInfo>
  <mods:name type="personal">
    <mods:namePart type="given">Rachel</mods:namePart>
    <mods:namePart type="family">Hardy</mods:namePart>
    <mods:role>
      <mods:roleTerm type="text">author</mods:roleTerm>
    </mods:role>
  </mods:name>
  <mods:name type="personal">
    <mods:namePart type="given">Charles</mods:namePart>
    <mods:namePart type="family">Oppenheim</mods:namePart>
    <mods:role>
      <mods:roleTerm type="text">author</mods:roleTerm>
    </mods:role>
  </mods:name>
  <mods:name type="personal">
    <mods:namePart type="given">Tim</mods:namePart>
    <mods:namePart type="family">Brody</mods:namePart>
    <mods:role>
      <mods:roleTerm type="text">author</mods:roleTerm>
    </mods:role>
  </mods:name>
  <mods:name type="personal">
```

```

<mods:namePart type="given">Steve</mods:namePart>
<mods:namePart type="family">Hitchcock</mods:namePart>
- <mods:role>
  <mods:roleTerm type="text">author</mods:roleTerm>
</mods:role>
</mods:name>
<mods:abstract>A primary objective of this research is to identify a framework for
universal citation services for open access (OA) materials, an ideal structure for
the collection and distribution of citation information and the main requirements of
such services. The work led to a recommended proposal that focuses on: • OA
contents in IRs rather than on wider OA sources. • Capture and validation of well-
structured reference metadata at the point of deposit in the IR. • Presentation of
this data to harvesting services for citation indexes. The aim of the proposal is to
increase the exposure of open access materials and their references to indexing
services, and to motivate new services by reducing setup costs. A combination of
distributed and automated tools, with some additional effort by authors, can be
used to provide more accurate, more comprehensive (and potentially free) citation
indices than currently exist.</mods:abstract>
- <mods:originInfo>
  <mods:dateIssued encoding="iso8601">2005-09</mods:dateIssued>
</mods:originInfo>
  <mods:genre>Other</mods:genre>
</mods:mods>

```

Obr. č. 4/6: metadatový záznam výzkumné zprávy [HARDY, 2005] z institucionálního repozitáře ECS v metadatovém formátu MODS [převzato 2008-12-12]

4.1.5 Digitální knihovna Nového Zélandu NZDL (část počítačové vědy)

4.1.5.1 Úvodní charakteristika systému

Zcela jiný model digitální knihovny, ve srovnání s předchozími americkými reprezentanty, je uplatněn v digitální knihovně NZDL (The New Zealand Digital Library, <http://www.nzdl.org/>, novější URL: <http://www.sadl.uleth.ca/nz/cgi-bin/library>). Její projekt vznikl v roce 1995 na Fakultě počítačové vědy (Department of Computer Science, <http://www.cs.waikato.ac.nz/>) **Waikatské university** (The University of Waikato, Hamilton, Nový Zéland, <http://www.waikato.ac.nz/>) za sponzorování několika organizací přímo na Novém Zélandu. V roce 1999 byl projekt sponzorován organizací FAO (OSN). Ačkoliv je pozornost v rámci tohoto textu zaměřena na digitální knihovny v oblasti počítačové vědy, je třeba podotknout, že NZDL je digitální knihovnou polytematickou (zahrnuje více než 40 různorodých digitálních fondů). Jeden z fondů patří také textům z **počítačové vědy** (jde o časově uzavřený fond), ale zahrnuje také například významné sbírky informací týkajících se Nového Zélandu jako země či regionu. Velmi zajímavé jsou specializované kolekce jiných typů digitalizovaných informací – obrazových, audiovizuálních a zvukových. Digitální knihovna zahrnuje i některé čistě bibliografické databáze importované ze zahraničních serverů.

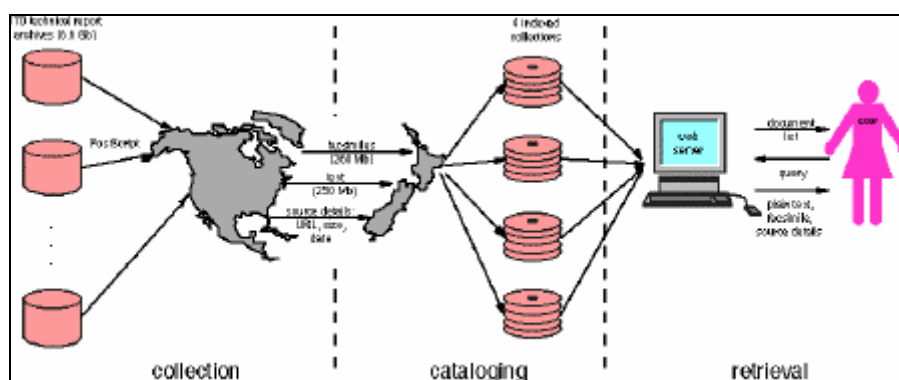
Pokud jde o samotnou oblast počítačové vědy, tak je v NZDL zastoupena dílčími digitálními fondy:

- **Human Computer Interaction (HCI) Bibliography:** jde o bibliografickou databázi záznamů textových dokumentů ze specifické oblasti počítačové vědy s relacemi do jiných oborů; databáze je zrcadlem původního amerického zdroje; volně dostupné texty jsou k záznamům připojovány (systém tak nabývá hybridního charakteru)
- **Computer Science Bibliographies:** jde o opět rozsáhlou bibliografickou databázi záznamů dokumentů z oblasti počítačové vědy, která je zrcadlem původního německého zdroje. Zahrnuje více než 1 000 000 záznamů

- **Computer Science Technical Reports (CSTR)**: jde o základní digitální fond úplných textů výzkumných zpráv z oboru počítačové vědy s celosvětovým záběrem. Zahrnuje více než 45 000 dokumentů, a patří tudíž současně chvíli k větším fondům v této disciplíně. Dokumenty pocházejí z FTP archivů nebo repozitářů významných severoamerických, asijských i evropských institucí (jde zhruba o 300 sídel); jde o uzavřený fond

4.1.5.2 Model digitální knihovny NZDL/CSTR (vstupy a jejich zpracování)

Schéma základních procesů budování a provozování digitálního fondu **CSTR** knihovny **NZDL** je znázorněno na obr. 4/7 (obrázek pochází z poloviny 90. let, nezachycuje již současný stav sběru dat). Dokumenty byly získávány z volně přístupných repozitářů a FTP archivů na serverech univerzit, resp. jiných výzkumných institucí Severní Ameriky, ale dnes i v dalších oblastech světa. Dokumenty z oblasti Severní Ameriky byly zjišťovány, identifikovány a stahovány automaticky v komprimované podobě (gzip) pomocí speciálních programů a transferovány na počítač partnerské organizace na Univerzitě v Calgary (Kanada) z důvodu předzpracování souborů.



Obr. 4/7: původní funkční schéma procesů digitální knihovny NZDL [WITTEN, 1996]

Na počítači v Calgary se ze stažených souborů extrahovaly jednak první a popřípadě i druhá stránka dokumentu jako obrazové faksimile, jednak úplný text celého dokumentu, který byl převáděn z formátu **PostScript** do jednoduchého řádkového formátu **ASCII** (pomocí programu PreScript). Původní stažené soubory ve formátu PostScript se vymazávaly. Takto upravené soubory byly na Nový Zéland přenášeny v nočním provozu.

Na obr. 4/7 je ve schématu dále vidět proces následné „katalogizace“ získaných dokumentů. Nejde však v žádném případě o klasickou ruční tvorbu záznamů dokumentů. V rámci přijatého modelu jde o zcela automatickou indexaci úplných textů dokumentů. Program, který zajišťoval toto zpracování, je označován zkratkou **MG (Managing Gigabytes)**. Program MG je součástí celého balíku programů, který nese označení **The Greenstone Software** (od novozélandské firmy DigiLib Systems Limited). V procesu zpracování byly jeho základním výstupem 4 dílčí indexy:

1. index úplných textů bez rozlišování velkých a malých písmen
2. index úplných textů s rozlišováním velkých a malých písmen
3. index individuálních stránek původního textu
4. index 1. stránky dokumentu

Úplné texty byly v rámci programu MG podrobeny řadě známých procedur pro zpracování úplného textu (lemmatizaci, vážení termínů, výpočet frekvence termínů aj.). Indexační soubory jsou uloženy na disku počítače v efektivní komprimované podobě. Na zpracování úplných textů se podílely i další specializované programy, např. **Sequitur** - program k identifikaci a rozpoznávání lexikálních struktur, nebo **Kea** - program k extrahování klíčových frází z textu.

4.1.5.3 Vyhledávání informací z digitální knihovny NZDL

Rozhraní pro vyhledávání z fondů NZDL má od roku 1999 nový design. Pokud jde o fond CSTR pro počítačovou vědu, je uplatněno plně vyhledávání v úplných textech dokumentů. Vyhledávat lze dvojím způsobem:

- v režimu **prohlížení** (browse), které vede buď k seznamu stručných záznamů dokumentů generovaných automaticky z indexů databáze úplných textů (viz příklad záznamu na obr. 4/8) nebo k seznamu odkazovaných FTP serverů, na nichž se nachází větší množství dokumentů k zadané problematice. Závěrečná volba získání úplného textu znamená stažení úplného textu zpravidla ve formátu PostScript z původního místa uložení. Pokud uživatel chce dokument ve formátu ASCII, je lepší použít způsob přímého vyhledávání, v jehož rámci lze získat úplný text v řádkovém ASCII formátu, který je lokalizovaný přímo na serveru Waikatské univerzity
- v režimu **přímého vyhledávání** je možné využívat buď jednoduchý způsob formulace dotazu (simple), který je standardní, nebo způsob pro pokročilé (advanced). V rámci pokročilého vyhledávání (jeho nastavení je nutné provést v rámci funkce Preferences) je možné vyhledávat na základě volně zapisovaných klíčových slov nebo frází (užívá se uvozovek) s tím, že řazení výsledků je nastaveno volbou „**ranked**“ (aplikace lokalizačně frekvenční funkce v úplném textu). Pokud je nastavena volba „**boolean**“ je možné využívat logických operátorů (AND = &, OR = | a NOT = !]. V rámci funkce Preferences je možné rovněž volit vyhledávání termínů s rozlišováním velkých a malých písmen (viz typy indexů popsaných výše). Nastavit lze rovněž hledání termínů s automatickým rozšířením zprava (dáno lemmatizací). Nalezená množina výsledků rešerše vede v první fázi k zobrazení seznamu záznamů.



MetaData for Database Mining John Cleary, Geoffrey Holmes, Sally Jo Cunningham, and Ian H. Witten Department of Computer Science University of Waikato Hamilton, New Zealand.

Abstract: At present, a machine learning application is accomplished by carefully crafting a single table from an often complex, multi-table database. The metadata necessary to create this table is rarely formally recorded, and is sometimes implicit in the structure of the database or the typing of the attributes. We categorize the types of metadata that we have encountered in our work with machine learning applications in agriculture, and describe a first generation tool that we have built to aid in the recording and use of metadata in database mining.

Obr. 4/8: příklad krátkého výstupního záznamu v digitální knihovně NZDL

V další fázi předkládá systém uživateli obrazovku s řadou dílčích funkcí týkajících se zobrazování úplného textu: je možné listování celým dokumentem po stránkách (se nebo bez zvýraznění hledaných slov), zobrazovat dokument v ASCII formátu nebo v původním PostScript formátu (downloading z původního FTP serveru) a zobrazování obrázků extrahovaných z dokumentu v průběhu zpracování. Ve zvláštním rámečku systém zobrazuje faksimile prvních stránek.

4.2 Systém digitální knihovny pro oblast ekonomiky RePEc

4.2.1 Úvodní charakteristika systému

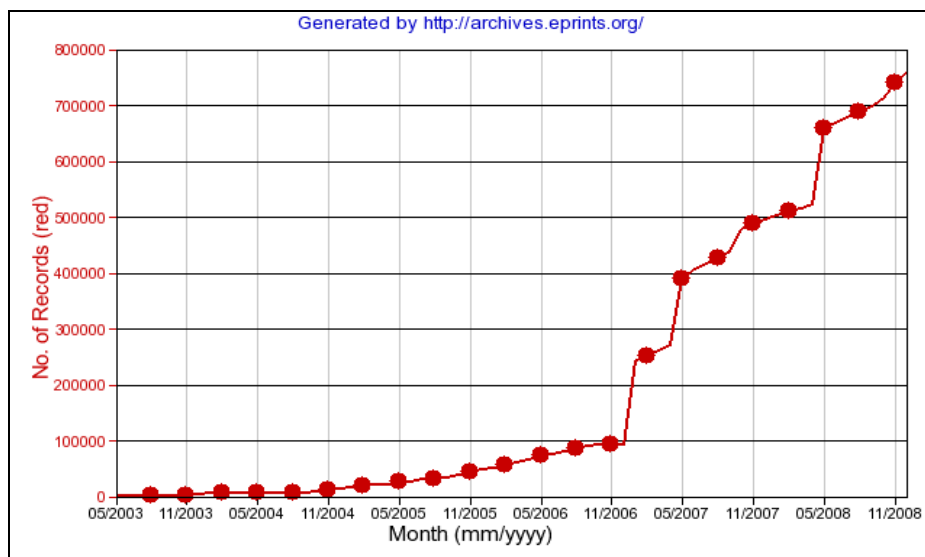
Významný mezinárodní systém distribuované digitální knihovny pro oblast ekonomiky **RePEc** (Research Papers in Economics, <http://repec.org/>) je jedním z největších systémů, který zpřístupňuje podstatnou část svého registrovaného obsahu v elektronické formě volně (zdarma) uživatelům z celého světa. Jde o nekomerční systém jehož cílem je podpora komunikace vědeckých informací z oblasti ekonomiky a příbuzných oborů. Systém je unikátní ve spojení **nepublikované a publikované literatury**. Ve světovém registru repozitářů ROAR zaujímá RePEc v současné době co do počtu záznamů 3. místo. Decentralizovaně budovaný systém registruje konci roku 2008 (dle údajů služby IDEAS) již cca **685 000 dokumentů**. Zhruba **571 000** z nich je dostupných jako úplný text zdarma (přístup k textům některých dokumentů, zejména článků z komerčních časopisů, je vázán na licenci). Registrované dokumenty reprezentují dokumenty z 64 zemí světa. Princip otevřeného přístupu k dokumentům je jedním ze zásadních principů v informační politice systému. Znamená to ale také, že náklady na provoz decentralizovaného systému jsou přeneseny zejména na provozovatele účastnických archivů (WWW, FTP), tedy instituce, které se dobrovolně do systému přihlásily, ale i na jednotlivce. Celý systém je zajišťován přibližně **100 dobrovolníky** (působí v lokálních archivech), kteří se starají zejména o to, aby systému vstupovala kvalitní vědecká data z daného oboru. Jinak se uplatňují jako správci lokálních archivů, techničtí konzultanti, editoři některých služeb, analytici kvality metadat aj.

Fungování celého systému bylo jeho čelními představiteli vyjádřeno výstižně následujícím způsobem [BARRUECO, 2000b, s. 230]:

mnoho archivů → jedna databáze → mnoho služeb

Celkový objem hlavní registrační databáze RePEc je úctyhodný a vybízí minimálně ke srovnání s objemem největší komerční mezinárodní bibliografické databáze z oblasti ekonomiky *EconLit* (<http://www.econlit.org/>), jejímž producentem je Americká ekonomická společnost (<http://www.aeaweb.org/>). V roce 2008 tato databáze obsahovala cca 850 000 záznamů, ovšem za retrospektivní období od roku 1969. Lze konstatovat v souladu s P. Jacsó, že databáze RePEC zahrnuje více dokumentů, plynoucích zejména z oblasti výzkumu (výzkumné zprávy, často užívaný angl. termín „working papers“) a navíc zahrnuje hyperlinky k úplným textům [JACSÓ, 2004b, s. 60]. Databáze *EconLit* je silná v registraci článků z komerčních vědeckých časopisů. V posledních 3 letech bylo zaznamenáno velké tempo růstu dokumentů registrovaných v databázi RePEc (viz obr. 4/9). Protože se producent databáze *EconLit* snaží registrovat i výzkumné zprávy (zařazuje již na základě smlouvy záznamy z databáze *AWPE* (Abstracts of Working Papers in Economics) vydavatelství Cambridge University Press), navázal spolupráci i se systémem RePEc s cílem poskytovat záznamy výzkumných zpráv či pracovních studií do databáze *EconLit* (včetně případného odkazu na úplný text). Systém RePEc má tak šanci být více vytěžován. V rámci dohody naopak systém RePEC využívá pro věcné pořádání dokumentů oborovou klasifikaci pro

ekonomiku *JEL*, kterou Americká ekonomická společnost používá jak v databázi *EconLit*, tak v její tištěné verzi – v časopise „*Journal of Economic Literature (JEL)*“. Vztah těchto dvou velkých mezinárodních zdrojů v oblasti ekonomické literatury je do budoucna otevřený, jednou z možností je i zánik některého z nich nebo nějaká forma sloučení. Systém RePEc díky optimálním podmínkám dalšího rozvoje informačních technologií i díky spolupráci s komerčními vydavateli časopisů může v budoucnu značně posílit registraci článků – může se stát hlavní registrační databází (vedle své hlavní role zpřístupňovat volně vědeckou literaturu z výzkumů).



Obr. 4/9: graf přírůstků registrační databáze systému RePEc za léta 2003-2008 [získáno z registru ROAR na základě dat agregátorské služby Celestial, 2008-12-22]

Zakladatelem systému RePEc je ekonom Thomas Krichel (t.č. působí na Knihovnické škole Univerzity na Long Islandu v USA), který má v systému i jako autor přidělen identifikátor s pořadovým **číslem 1** (RePEc Short-ID: pkr1) (<http://ideas.repec.org/e/pkr1.html>). Má několik nejbližších spolupracovníků, kteří na mají na starosti zejména správu specifických služeb a subsystémů. Systém RePEc vznikl na základě původního distribuovaného projektu systému WoPEc (Working Papers in Economics) [BARRUECO, 2000b, s. 229], finančně byl podpořen granty z Velké Británie v rámci programu *E-Lib*. Systém byl zajišťován konsorciem Internetového projektu pro akademické ekonomy pod názvem *NetEc* (jeho činnost byla ukončena v březnu 2005). Data původní databáze WoPEc se později stala součástí jedné ze současných služeb systému RePEc - služby EconPapers (<http://econpapers.repec.org/>). V květnu 1997 pak byla v rámci rozrůstající se sítě spolupracujících archivů založena současná registrační a dokumentační databáze RePEc (<http://repec.org/>) pro zpracování a zpřístupňování různých typů vědeckých dokumentů z oblasti ekonomiky, i když to tradiční název systému nevyjadřuje přesně.

Systém RePEc, který je v již zmíněném registru ROAR zařazen pod kategorii „Výzkum napříč institucemi“ (Research Cross-Institutional), využívá pro svoji hlavní bibliografickou databázi i pro svoje subsystémy a služby **různé softwarové aplikace**. V registru ROAR je registrován od 13. května 2003, od kdy jsou mu také měřeny statistiky přírůstků (viz obr. 4/9). Systém je také plně připraven na sběr metadat podle protokolu OAI-PMH (<http://oai.repec.openlib.org>).

Jádrem systému RePEc je jednak jeho **bibliografická databáze** registrující zejména pracovní studie a výzkumné zprávy (šedá literatura), články z časopisů a softwarové komponenty, jednak velké množství jeho **specifických služeb a subsystémů** (viz dále v textu).

4.2.2 Účastníci systému RePEc

Účast v systému RePEc je zcela **dobrovolná**. Členství je otevřené pro každou instituci, která má zájem spolupracovat na vytváření této globální digitální knihovny. Členské instituce pak mají právo vytvářet podle stanovené metodiky systému RePEc metadatové záznamy dokumentů, které vlastní a mají uložené v lokálním archivu (institucionálním repozitáři nebo archivu jiného typu). V současné chvíli jsou data stahována z cca **900 lokálních archivů**. Metadata jsou pak pomocí specifického protokolu stahována do centrální metadatové databáze. Pokud nějaká instituce není členem systému, může uživatel ukládat svoje příspěvky do systému RePEc prostřednictvím mezinárodního archivu MPRA (Munich Personal RePEc Archive). Uložená metadata jsou rovněž stahována do centrální registrační databáze. Systém RePEc nepodporuje vytváření osobních archivů. Registr institucí zahrnoval ke konci roku 2008 cca **10 800 institucionálních členů**.

Mezi nejvýznamnější členy systému RePEc patří například nakladatelské organizace, jako je společnost Elsevier, Wiley Blackwell, Springer, Oxford University Press, Cambridge University Press, MIT Press, dále již výše uvedená Americká ekonomická asociace, Mezinárodní měnový fond, Světová banka, řada redakcí ekonomických časopisů, dílčích součástí velkých univerzit a jiných akademických institucí. Komerční vydavatelé prostřednictvím svých digitálních knihoven zajišťují pro systém RePEc celkem 2 700 specializačních řad.

4.2.3 Typy zahrnovaných dokumentů a nástroje pro zpracování metadat

Do databáze systému RePEc přijímá následující typy informačních dokumentů (v závorkách je uveden jejich počet ke konci roku 2008 podle údajů služby IDEAS):

- **články z časopisů a jiných periodik** (400 159, z toho 347 060 volně online dostupných)
- **pracovní studie (working papers)** (278 069, z toho 114 446 volně online dostupných)
- **kapitol z knižních publikací** (2 795, z toho 2 661 volně online dostupných)
- **knižní publikace** (2 344, z toho 1 290 volně online dostupných)
- **softwarové komponenty** (1 682, toho 1 668 volně online dostupných)

Systém dále v rámci celé koncepce zpracování metadat vede také záznamy excerpovaných **periodik a ekonomických edic či jejich řad** (knižních titulů, pracovních studií, je jich v současné době 3 467) a také záznamy registrovaných **účastnických archivů** (je jich, jak bylo výše uvedeno, cca 900).

Distribuovaná architektura systému RePEc znamená jednak vytváření metadatových záznamů stanovených objektů na základě vlastní metadatové specifikace **ReDIF** (Research Documents Information Format) [KRICHEL, 2007], připravené na základě definic známých formulářů IAFA (IAFA templates), a následně jejich sklizení do centrální metadatové báze na základě vlastního komunikačního protokolu zvaného „**Guildford protocol**“ (**GuilP**) [KRICHEL, 2007].

Metadata se vyplňují podle stanoveného návodu systému RePEc. K dispozici jsou **předefinované formuláře pro jednotlivé typy objektů**:

- Template-Type: ReDIF-Archive 1.0 for archive templates (archiv)

- Template-Type: ReDIF-Series 1.0 for series templates (edice/seriál)
- Template-Type: ReDIF-Paper 1.0 for working paper templates (pracovní studie)
- Template-Type: ReDIF-Article 1.0 for journal article templates (článek)
- Template-Type: ReDIF-Book 1.0 for book templates (kniha)
- Template-Type: ReDIF-Chapter 1.0 for book chapter templates (kapitola)
- Template-Type: ReDIF-Software 1.0 for software component templates (softw. komponenta)

Jednotlivé formuláře zahrnují povinná a nepovinná metadata [BARRUECO, 2005, s. 92]. Vybraný příklad metadatového záznamu článku z časopisu uveden na obr. 4/10 (povinné údaje jsou v příkladu uvedeny kurzivou). Hlavním povinným údajem je jedinečný identifikátor „**Handle**“, který je uveden v posledním poli (skládá se z kódů reprezentujících řadu významových informací, například o ročníku časopisu, o čísle sešitu, stránkování aj.). Identifikuje daný objekt v celém systému RePEc. Pro získání úplných textů je důležité uvádět URL adres. Metadatové soubory jsou jednoduché textové soubory a jsou komunikovány pomocí výše jmenovaného protokolu jako soubory s příponou „**.rdf**“. Systém využívá, jak bylo již výše v textu zmíněno oborové klasifikace JEL Americké ekonomické asociace.

```

Template-Type: ReDIF-Article 1.0
Author-Name: Andrew Beacon
Author-Name: Charlie Doe
Author-Email: doe@someuni.edu
Author-Workplace-Name: Some Big University
Author-Name: Engelbert Falafel
Author-Workplace-Name: Some Other University
Title: A Guide to Forecasting Agricultural Output by Following
Weather
Patterns in Neighboring Areas
Abstract: In this article, we survey the literature on agricultural
output forecasting and suggest some avenues the merit particular
attention...
Classification-JEL: R00, Z0
Keywords: Forecasting, agricultural economics
Journal: Annals of Computational Economics
Pages: 1-58
Volume: 3
Issue: 1
Year: 1999
Month: February
File-URL: http://www.geekeconomics.com/articles/thefile.pdf
File-Format: Application/pdf
File-URL: http://www.geekeconomics.com/articles/thefile.html
File-Format: text/html
Handle: RePEc:aaa:journl:v:3:y:1999:i:1:p:1-58

```

Obr. 4/10: vstupní metadatový záznam článku z časopisu v systému RePEc [získáno z WWW dokumentace systému 2008-12-22]

4.2.4 Služby systému RePEc

V této podkapitole budou v dílčích částech představeny jednotlivé služby nebo subsystémy celého systému RePEc včetně „služby“ IDEAS pro využívání hlavní bibliografické (metadatové) databáze systému. Komentovány podrobněji ale nebudou příbuzné služby

(zajišťované jinými partnerskými subjekty), které systém RePEc odkazuje z hlavní webové stránky. Jde o služby:

- **Inomics** (<http://www.inomics.com/cgi/show>) nabízí stahování textů výzkumných zpráv či pracovních studií společně s informacemi o ekonomických konferencích, pracovních příležitostech a také specifické vyhledávání v internetu pro ekonomy
- **Socionet Personal Zone** (<http://spz.socionet.ru/index-en.shtml>) je službou společného pracovního prostoru, která je určena pro badatele, učitele a studenty v rámci informačního prostoru systému RePEc (služba je zajišťována ruským subjektem)
- **Socionet (Соционет)** (<http://socionet.ru/>) je ruskou (v ruském jazyce) implementací metod a databáze systému RePEc pro informační prostředí sociálních věd

Personal Details

First Name: Dominik
Middle Name:
Last Name: Vymetal
Suffix:
RePEc Short-ID: pvy5
Email: [This author has chosen not to make the email address public]
Homepage:
Postal Address:
Phone:

Affiliation

(in no particular order)

- **Obchodně Podnikatelská Fakulta (School of Business Administration)**
Slezská Univerzita (Silesian University)

Location: Opava, Czech Republic
Homepage: <http://www.opf.slu.cz/>
Email:
Phone: 00420 596 398 111
Fax: 00420 596 312 069
Postal: Univerzitní nám. 1934/3, 733 40 Karviná
Handle: RePEc:edi:opfslez ([registered authors at this institution](#))

Works

[Working papers](#) | [Access and download statistics](#) | [Citations](#) (if any) | [NEP Fields](#) |
Download all references for this author: available formats: [HTML](#), [plain text](#), [BibTeX](#), [RIS](#) (EndNote), [ReDIF](#)

Working papers

Obr. 4/11: autoritní záznam českého vědce (Slezská univerzita, Opava) s identifikátorem (pvy5) prezentovaný v rámci služby IDEAS systému RePEc [převzato z WWW 2008-12-22]

4.2.4.1 Registrace autorů v rámci Autorské služby RePEc

Dva subsystémy zabezpečují pro RePEc jedinečné identifikační služby. První z nich je **Autorská služba RePEc** (RePEc Author Service, <http://authors.repec.org>), v jejímž rámci jsou spravovány potřebné informace (údaje) o autorech přispívajících do systému RePEc. V současnosti (ke konci roku 2008) je v systému registrováno přibližně **18 543 autorů** (též autorů z ČR). V tomto systému je také podle posledních údajů zaregistrováno dalších cca 6 000 fyzických osob, které zatím nemají na kontě žádné autorství v systému RePEc. Službu řídí T. Krichel a databáze registru je provozována na serveru Ekonomické fakulty

Connecticutské univerzity, USA (sponzoring zajišťuje Společnost pro ekonomickou dynamiku, Society for Economic Dynamics). V rámci této služby se musí registrovat každý nový autor systému RePEC. Sám vyplňuje pomocí formuláře řadu údajů přes webové rozhraní (jméno, varianty jména, afilační údaje, e-mailová adresa aj.). Systém pak na základě těchto informací pro autora zajišťuje propojování autoritního záznamu se záznamy jeho dokumentů. V rámci registračního procesu je autorům přidělován **jedinečný identifikátor**, který se skládá z vybraných znaků jména autora a nějakého významného údaje typu data. Krátká forma identifikátoru se pak zobrazuje také v prezentačním záznamu v systému IDEAS (<http://ideas.repec.org/e/pvy5.html>), v databázi NEP aj.

Na obr. č. 4/11 je prezentační formát záznamu autora z České republiky, jehož zkrácený identifikátor je:

RePEc Short-ID: pvy5

V rámci databáze služby IDEAS je možné **vyhledávat záznamy autorů** (i jejich propojených prací) pomocí rejstříků podle řady vybraných hledisek: abecedně podle jmen autorů, podle afilace, podle oborové specializace (zatřídění autorů je realizováno automaticky na základě služby NEP – viz dále). K dispozici jsou také specifické seznamy nově zařazených autorů, autorů – držitelů Nobelovy ceny za ekonomii, špičkoví ekonomové podle citovanosti aj.

Vysoká Škola Ekonomická v Praze

Translation: University of Economics
Location: Praha, [Czech Republic](#)
Homepage: <http://www.vse.cz/>
Email:
Phone: (02) 24 09 51 11
Fax: (02) 24 22 06 57
Postal: nam. W. Churchilla 4, 130 67 Praha 3
[Handle](#): RePEc:edi:uevsecz
[Areas or Functions](#):

Subentities listed with EDIRC:

1. [Fakulty Financí a Účetnictví, Vysoká Škola Ekonomická v Praze](#)
2. [Katedra Bankovníctví a Pojišťovníctví, Fakulty Financí a Účetnictví, Vysoká Škola Ekonomická v Praze](#)

Additional details:

People who have registered with **RePEc** and have claimed to be affiliated with this institution:

NEW! Find all publications of all authors listed below conveniently compiled [on IDEAS](#).

1. [Adamovic, Jan](#)
2. [Buus, Tomas](#)
3. [Janda, Karel](#)
4. [Klazar, Stanislav](#)
5. [Kotlán, Viktor](#)
6. [Sindel, Jaromir](#)
7. [Witzany, Jiri](#)

The screenshot shows a web browser window with a taskbar at the bottom indicating 'Internet' and '100%' zoom.

Obr. 4/12: záznam české univerzity včetně identifikátoru v databázi subsystému EDIRC systému RePEc (edi:uevsecz) [převzato z WWW 2008-12-22]

4.2.4.2 Registrace institucí v rámci služby EDIRC

Druhý registrační subsystém, známý pod zkratkou **EDIRC** (Economics Departments, Institutes and Research Centers in the World, <http://edirc.repec.org>), zajišťuje správu autoritních informací o všech institucích (korporacích), ze kterých afiličně pocházejí autoři systému RePEc nebo dalších spolupracujících organizací. V současnosti je v databázi této služby registrováno přibližně **11 017 institucí** z cca **229 zemí** (ekonomické ústavy nebo katedry univerzit, výzkumná centra, ministerstva financí, statistické úřady, banky aj.). Správcem této služby je Christian Zimmerman a databáze registru je provozována rovněž na serveru Ekonomické fakulty Connecticutské univerzity, USA. Jednotlivých korporacím je v autoritním záznamu přidělován také **jedinečný identifikátor**. Každý identifikátor korporace obsahuje celkem 7 znaků (<http://edirc.repec.org/handle.html>): **RePEc:edi:iiiiic**

Prvních pět znaků označuje instituci, další dva znaky reprezentují zemi či stát (užívá se kódů ISO pro internetové domény, například ekonomické asociace mají kód „ea“:

RePEc:edi:aeaaaaea (American Economic Association)

Příklady přidělených identifikátorů některých institucí z ČR (z rejstříku služby EDIRC, <http://edirc.repec.org/alphabet.html>):

RePEc:edi:uevsecz (*Vysoká škola ekonomická, VŠE*) – viz obr. 4/12

RePEc:edi:ffvsecz (VŠE, Fakulta Financí a Účetnictví, FFÚ)

RePEc:edi:dbvsecz (VŠE, FFÚ, Katedra bankovníctví a pojišťovnictví)

Služba EDIRC nabízí **vyhledávání záznamů institucí** (a propojených prací) pomocí několika rejstříků podle řady vybraných hledisek: abecedně podle jmen zemí, podle zemí v rámci kontinentů, podle zemí na základě kódu země a podle oborové specializace institucí. K dispozici je také rozhraní pro dotazové vyhledávání záznamů (Search, uplatnit lze klíčová slova z polí URL, jméno instituce, překlad jména instituce, místo sídla instituce). Nabízen je také specifický rejstřík ekonomických společností a asociací.

4.2.4.3 Služba IDEAS a registrační databáze systému RePEc

Služba **IDEAS** (Internet Documents in Economics Access Service, <http://ideas.repec.org/>) poskytuje uživatelům z celého světa informace na základě hlavní bibliografické (metadatové) databáze systému digitální knihovny RePEc. Databáze je provozována také na serveru Ekonomické fakulty Connecticutské univerzity (USA) pod vedením Christiana Zimmermanna. Sponzorem služby je již výše zmíněná Společnost pro ekonomickou dynamiku (Society for Economic Dynamics). Na správě databáze se podílí řada dalších odborníků z celého světa.

Databáze využívaná službou IDEAS má registrační ráz, to znamená, že zahrnuje také záznamy dokumentů, ke kterým uživatelé nemají volný přístup (je vázán na licenci). V současné době se v databázi nachází celkem **685 000 záznamů**, z toho k **571 000** z nich je volný online přístup. Typologie zastoupených dokumentů byla představena výše v textu.

Služba nabízí více **způsobů vyhledávání informací** (viz obr. 4/13). **Dotazové vyhledávání** lze realizovat ve dvou režimech, jednoduchém (Search), které je k dispozici ve formě jednoduchého okna z hlavní stránky služby (viz obr. 4/13, nebo pokročilem vícehlediskovém (more detailed). Uplatnit lze běžné selekční údaje na základě ukládaných metadat (jméno autora, jméno instituce aj.). Dotazovací jazyk umožňuje užití základních booleovských operátorů. Uživatel si též může zvolit i formát prezentovaného záznamu (krátký, dlouhý), lze také volit způsob seřazení záznamů (podle relevance, podle abecedy názvů dokumentu aj.).

Vyhledávat lze i **prohlížením rejstříku** podle typu dokumentu, podle klasifikace JEL aj.). Dotazové vyhledávání i prohlížení rejstříku se týká také záznamů autorů a institucí. Rešerše lze limitovat parametrem dostupnosti dokumentů volně a online.

V případě zájmu si lze pro potřeby vyhledávání nainstalovat speciální vyhledávací **plug-in**, který je nabízen z hlavní stránky služby IDEAS - dostupný je však jen pro prohlížeč Firefox.

Welcome to the largest bibliographic database dedicated to Economics and available freely on the Internet. Over 650'000 items of research can be browsed or searched, and over 550'000 can be downloaded in full text! This site is part of a large volunteer effort to enhance the free dissemination of research in Economics, **RePEc**. To see the popularity of these services, browse the statistics at [LogEc](#)

Authors: register with RePEc

Then, anyone finding some of your research here can find your latest contact details and a listing of your other research. You will also receive a monthly mailing about the popularity of your works, your ranking and new citations found.

Search [More detailed search](#)

IDEAS currently serves information about:

	Journals & series	Items	of which				
			online	JEL coded	with abstracts	with references	with citations
Working papers	2440	278069	219246	114466	211049	128482	99453
Articles	847	400159	347060	19150	156337	42751	121960
Chapters	54	2795	2661	1192	1014	128	682
Books	104	2344	1290	776	841	8	120
Software components	22	1682	1668	25	1577	48	125
Total	3467	685049	571925	135609	370818	171417	222340

In addition, IDEAS has information about
11017 economics institutions,
18543 authors who have registered with the **RePEc Author Service** (who have authored 371544 items listed in RePEc) and

Navigation links: [working papers](#), [journals](#), [software](#), [chapters](#), [books](#), [authors](#), [institutions](#), [Search](#), [Browser search plug-in](#), [JEL classification](#), [NEP reports](#), [Pub compilations](#) **NEW**, [Reading lists](#) **NEW**

Obr. 4/13: hlavní prezentační stránka databáze IDEAS s hlavní statistikou stavu a s nabídkou částí vyhledávacích funkcí [získáno z WWW 2008-12-22]

Z hlavní stránky služby IDEAS je možné pomocí příslušného hyperlinku přejít k vyhledávání informací poskytovaných další službou systému **NEP Reports** (New Economics Papers) – jde o službu zasílání záznamů o nových pracovních studiích pomocí-mailu (v 81 různých zájmových oblastech ekonomiky).

Služba IDEAS dále nabízí novou službu pro **sestavování uspořádaných seznamů** publikací podle parametrů vybraných samotným uživatelem (Pub Compilations). Nová je i služba **Doporučená literatura** (Reading list), díky které si uživatelé mohou sestavit svůj vlastní čtenářský seznam, který mohou později využít např. při výuce, při psaní studie na specifické téma aj. Tato služba umožňuje připravit souhrn veškeré potřebné relevantní literatury. Stejně jako ostatní části systému RePEc, i služba IDEAS poskytuje odkaz na blog RePEc, který poskytuje informace o různých novinkách v systému a diskuse na různá témata.

4.2.4.4 Mnichovský personální archiv RePEc (MPRA)

Digitální personální archiv **MPRA** (Munich Personal RePEc Archive, <http://mpra.repec.org/>) je součástí systému RePEc. Byl vytvořen s cílem sloužit autorům - ekonomům, kteří chtějí vystavovat svoje práce volně prostřednictvím systému RePEc, ale nejsou formálně členy žádné instituce, která by jim poskytovala podporu a tím institucionální členství v systému. Práce těchto autorů jsou pak zpřístupněny také prostřednictvím služby EconPapers, IDEAS aj. Archiv MPRA je řízen na základě dohody se systémem RePEc Univerzitní knihovnou v Mnichově (Universitätsbibliothek München, Ludwig-Maximilians-Universität München, <http://www.ub.uni-muenchen.de/>).

Repozitář MPRA přijímá k uložení zejména pracovní studie (working papers) nebo výzkumné zprávy, ale a knihy ve všech jazycích, které se týkají výzkumu v oblasti ekonomiky. Na obr. 4/14 je záznam práce českého odborníka v češtině (jeho autoritní záznam je na obr 4/11), uložený v roce 2008. MPRA také přijímá elektronické verze již dříve publikovaných článků (s podmínkou souhlasu vlastníka práv copyrightu). O zařazení elektronicky podaných dokumentů rozhoduje okolo 40 editorů z celého světa. Ke konci roku 2008 bylo v repozitáři archivu MPRA uloženo cca 6 510 dokumentů.

Archiv MPRA je provozován v softwaru aplikaci Eprints, jedné z nejnovějších verzí EPrints 3.0.5 (Puffin) z 21 dubna 2008. Ta nabízí sofistikované rozhraní pro ukládání metadat a úplných textů, řadu rozhraní pro vyhledávání informací dotazového typu i vyhledávání pomocí prohlížení četných rejstříků (pomocí jména autora, jména instituce, klasifikace JEL, jazyka aj.).

MPRA Munich Personal RePEc Archive

Home | Browse | Search | About | Help
Login | Create Account

Některé právní otázky související se zaváděním informačních systémů v nadnárodních firmách

Vymetal, Dominik (2008): *Některé právní otázky související se zaváděním informačních systémů v nadnárodních firmách*. Published.

Full text available as:

PDF - Requires a PDF viewer such as [GSview](#), [Xpdf](#) or [Adobe Acrobat Reader](#)
274Kb

Abstract

Abstract: One of noticeable features of globalization are rapid changes of information system carry out many actions having impact on information system. This increasingly concerns multi-national ERP system implementations. Some basic formal prerequisites must be fulfilled. Harmonized legal, currency and custom systems in target countries are to be mentioned here and culture are the next important point. After the delivering party is chosen, other important shall be taken into account in order to avoid conflicts, delays and language barriers. Proper help to solve eventual problems during project and implementation stage. These topics must managers, too. This is why it is an opposite praxis to include tuition of special law lectures in education.

Item Type: MPRA Paper

English Title: Some juridical questions related to information systems implementation in multi-national companies

Language: Czech

Keywords: Information system project; ERP; contract law; multi-national companies; delivery contract; multi-national solution team; IT project risks; bankruptcy of delivering party

Subjects: [F - International Economics > F2 - International Factor Movements and International Business > F23 - Multinational Firms; International Business](#)
[K - Law and Economics > K1 - Basic Areas of Law > K12 - Contract Law](#)
[L - Industrial Organization > L2 - Firm Objectives, Organization, and Behavior > L22 - Firm Organization](#)

http://mpra.ub.uni-muenchen.de/12406/1/MPRA_paper_12406.pdf

Obr. 4/14: část úplného záznamu zprávy českého autora uložené v repozitáři archivu MPRA v softwaru EPrints s náhledem titulní strany [získáno z WWW 2008-12-22]

Archiv MPRA nabízí služby typu RSS (1.0, 2.0) nebo ATOM pro rozesílání novinkových obsahů repozitáře. Tato služba může být uložena mezi záložky a později tak rychle vyvolána.

4.2.4.5 EconPapers

Cílem nadstavbové služby **EconPapers** (<http://econpapers.repec.org/>) je poskytovat rychlý přístup k **novým, online dostupným** ekonomickým pracovním studiím, článkům z časopisů, softwarovým komponentám aj.. Tato služba byla v době zrodu systému nazývána, jak bylo uvedeno výše v textu, WoPec (Working Papers in Economics).

Služba EconPapers využívá bibliografických dat registrační databáze systému RePEc. Je zajišťována Sunem Karlssonem, prof. statistiky na Fakultě obchodu, ekonomiky, statistiky a informatiky Univerzity v Örebro ve Švédsku (spravuje také švédský archiv pracovních studií SWoPec (<http://swopec.hhs.se>)). Služba je provozována na serveru této instituce. Většina textů dokumentů je dostupná zdarma, avšak některé dokumenty (zejména články z časopisů) vyžadují, aby organizace, v níž uživatel působí, byla registrována v rámci služby EconPapers.

V rámci služby EconPapers, stejně jako i v jiných službách systému RePEC, si uživatel může vybrat, zda bude vyhledávat **dokumenty typu** pracovní studie, články z časopisů, knihy nebo kapitoly z knih nebo softwarové komponenty. Na obr 4/15 je zobrazen záznam článku z online časopisu českých autorů s možností stažení jeho úplného textu z repozitáře Fakulty sociálních věd UK v Praze.

EconPapers
Economics at your fingertips

[EconPapers Home](#)
[About EconPapers](#)

[Working Papers](#)
[Journal Articles](#)
[Books and Chapters](#)
[Software Components](#)

[Authors](#)

[JEL codes](#)
[New Economics Papers](#)

[Advanced Search](#)

Quick Search

[EconPapers FAQ](#)
[Cookies at EconPapers](#)
[Archive maintainers FAQ](#)

[Format for printing](#)

[The RePEc blog](#)

RePEc
This site is part of RePEc and all the data displayed here is part of the RePEc data set.
Is your work missing from RePEc? Here is how to [contribute](#).

Hits and Misses: Ten Years of Czech Inflation Targeting (Introduction)

Kateřina Šmídková (katerina.smidkova@cnb.cz), Aleš Bulíř and Martin Čihák (mcihak@imf.org)
[Additional contact information](#)

Czech Journal of Economics and Finance (Finance a uver), 2008, vol. 58, issue 9-10, pages 398-405

Abstract: The first 10 years of Czech Republic's inflation targeting regime have been remarkable by the persistent undershooting of the inflation targets. This article is an introduction to a special issue aimed at analyzing the factors of the undershooting. The articles in this issue explore the following hypotheses: the impact of a series of counter-inflationary shocks; a biased forecasting system; a faulty decision-making system; and the role of inflation expectations. Based on a "meta-analysis" of the articles, this introduction suggests that the undershooting cannot be explained by a single factor, although counter-inflationary shocks are identified most frequently. The impact of the undershooting on central bank credibility was modest, partly because the forecasts were accompanied by explicit discussions of the forecast risks.

Keywords: [inflation targeting](#); [forecasting system](#); [inflation expectations](#) (search for similar items in EconPapers)

JEL-codes: [E52](#) [E58](#) [E31](#) [E47](#) (search for similar items in EconPapers)

Date: 2008

Downloads: (external link)
<http://journal.fsv.cuni.cz/mag/article/show/id/1139> (application/pdf)

Related works:
This item may be available elsewhere in EconPapers: [Search](#) for items with the same title.

[Access Statistics](#) for this article

[More articles](#) in *Czech Journal of Economics and Finance (Finance a uver)* from [Charles University Prague, Faculty of Social Sciences](#)
[Contact information at EDIRC](#)

Obr. 4/15: část úplného záznamu článku z online dostupného časopisu českých autorů (z roku 2008), zobrazeném v rámci služby EconPapers [získáno z WWW 2008-12-22]

V rámci služby EconPaper lze využít komfortní rozhraní pro **dotazové vyhledávání** jednoduchého typu (Search) i pokročilého vyhledávání (Advanced Search). **Jednoduché vyhledávání** zajišťuje rešerši provedenou nad všemi druhy dokumentů. Nabízeno je také specifické rozhraní historie rešerší a v rámci něho možnost jejich kombinování pomocí základních booleovských operátorů. Uživatel může volit uspořádání výsledných záznamů podle relevance nebo data uložení či aktualizace záznamu. **Pokročilé vyhledávání** nabízí užití více selekčních údajů najednou (jméno autora, klíčové slovo, název dokumentu, klasifikace JEL, typ dokumentu jako limitační údaj), umožňuje zpřesnění dotazu (slova – fráze). Stran softwarových komponent lze volit i selekční údaj druhu programovacího jazyka (Fortran, Gauss, Mathematica, Perl, Python, Rats, Shazam, Stata aj.).

Ze stránek služby EconPapers se může uživatel pomocí odkazů přesunout do dalších služeb systému RePEc - služby CitEc, NEP, IDEAS a LogEc.

4.2.4.6 Služba NEP

Služba **NEP** (New Economics Papers, <http://nep.repec.org/>) je službou automatického upozorňování na nové přírůstky v databázi RePEc. Využívá se e-mailového kontaktu s uživateli. Je možné využít také službu RSS. Služba je sponzorována Státní Newyorskou univerzitou (State University of New York at Oswego). Generování zpráv mají na starosti předmětově erudovaní editoři.

Registrace k odběru služby je zdarma a je dostupná bez omezení. V současné době je možné si vybrat z 83 předmětových zájmových oblastí – ty jsou označeny specifickým kódem (například NEP-AFR pro dokumenty týkající se Afriky, NEP-PUB pro dokumenty týkající se veřejných financí). Zvolit lze také odběr seznamu všech nových přírůstků.

Služba NEP je přístupná jak přes systém IDEAS tak i prostřednictvím EconPapers.

K registraci ke službě NEP stačí jen aktivní e-mailová adresa, uživatel si pak musí zvolit přístupové heslo. Jméno uživatele není vyžadováno. Uživatel si může také zvolit jazyk, ve kterém mu budou zprávy posílány (jen výběr jazyků). Registraci je nutné potvrdit přes automatický e-mail, který je zájemci poslán. V současnosti systém zajišťuje **50 257** „předplatných“ k této službě, zprávy jsou zasílány na 22 856 adres z celého světa.

4.2.4.7 LogEc

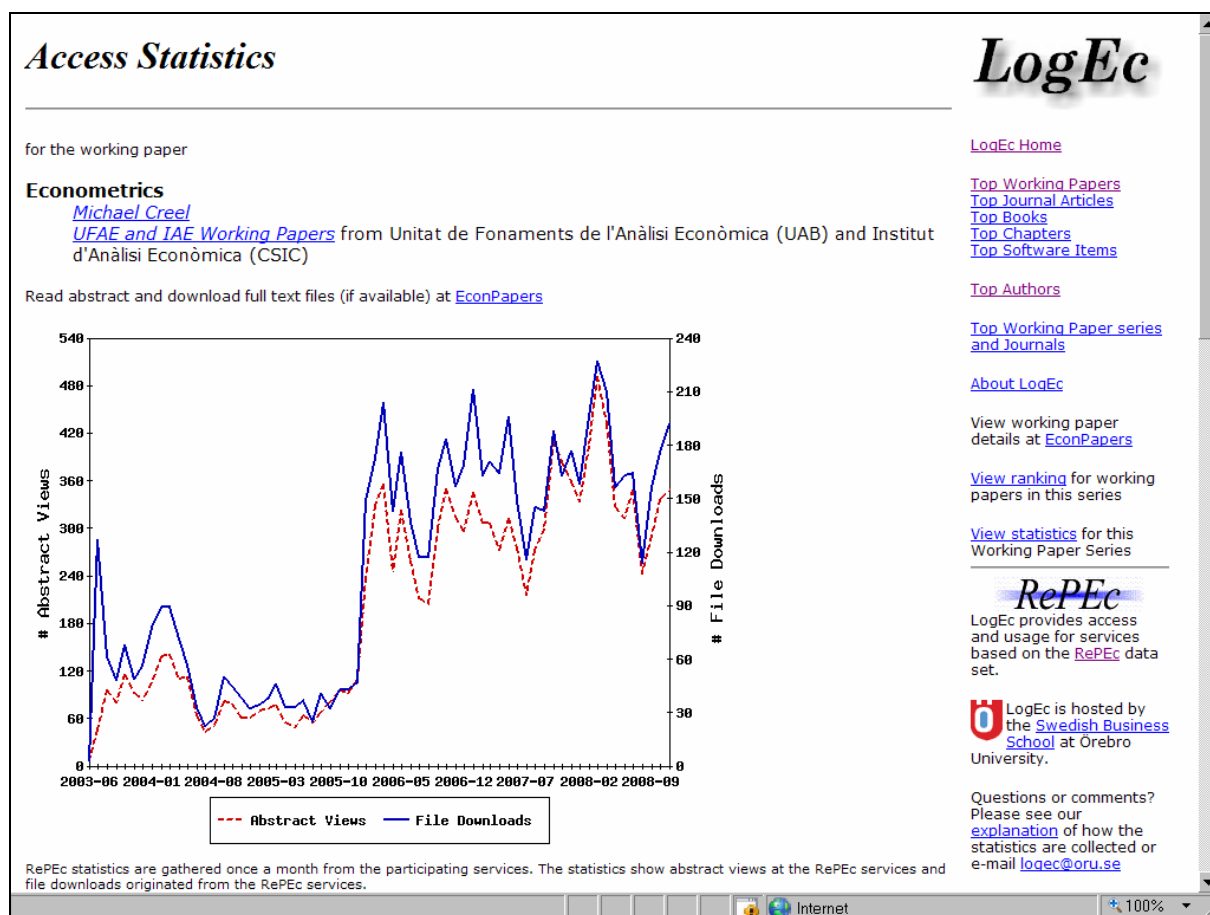
Služba **LogEc** (<http://logec.repec.org/>) poskytuje zdarma uživatelům cenné statistické údaje týkající se **využívání** celého systému RePEc, zejména pak jeho dílčích částí, které vycházejí z ústřední bibliografické databáze (služby EconPapers, IDEAS, NEP, ale i Socionet), jakož i každého dokumentu v digitální knihovně zahrnutého. Je zajišťována opět již zmíněným S. Karlssonem a je provozována na serveru Fakulty obchodu, ekonomiky, statistiky a informatiky Univerzity v Örebro ve Švédsku. Podle ohlasů uživatelů, jde o vynikající a velmi sledovanou službu pro ekonomy [JACSÓ, 2004b, s. 58].

Statistiky jsou aktualizovány zhruba jednou měsíčně, kdy jsou sebrána a sjednocena data a soubory logů ze serverů spolupracujících organizací. Aby byly statistiky co nejpřesnější a odpovídaly skutečným přístupům, jsou ze záznamů vyňaty přístupy automatů a robotů (programy, které automaticky a systematicky prohledávají internet, vytvářejí indexy - mohou tvořit až 60 % přístupů), a také dvojí či vícenásobné přístupy jednoho uživatele (kdy uživatel stáhne jeden dokument dvakrát).

Služba zobrazuje celkový počet stažených souborů a zobrazených abstraktů za konkrétní měsíc. Zároveň zobrazuje statistiku od začátku provozu systému RePEc v lednu 1998 (celkem 35 609 309 stažených souborů a 156 449 170 zobrazených abstraktů ke konci roku 2008).

V grafické podobě nabízí rozlišení zobrazení abstraktů a stažení souborů (viz obr. 4/16) podle jednotlivých služeb (EconPapers, IDEAS, NEP a Socionet). Dalším rozlišením, které LogEc znázorňuje graficky, je typ dokumentu, nejčastěji stahovanými dokumenty a zobrazovanými abstrakty jsou pracovní studie (working papers). Služba zobrazuje také celkový graf počtu záznamů v databázi RePEc.

Kromě grafického vyjádření je možné si souhrnné statistiky zobrazit i ve formě textové tabulce (není příliš přehledná).



Obr. 4/16: graf statistiky přístupu k vybranému dokumentu systému RePEC zajišťovaný službou LogEc [získáno z WWW 2008-12-22]

Dále služba LogEc nabízí prohlížení seznamu nejčastěji zobrazovaných záznamů v posledním měsíci. Počet zobrazených záznamů je možné systémem určit. Údaje v tomto žebříčku jsou rozděleny podle počtu stažených dokumentů a zobrazených abstraktů, které se k danému záznamu vztahují, a kromě posledního měsíce prezentují i počty zobrazení za tři předchozí měsíce a celkem. Podle těchto kritérií je možné zobrazování filtrovat.

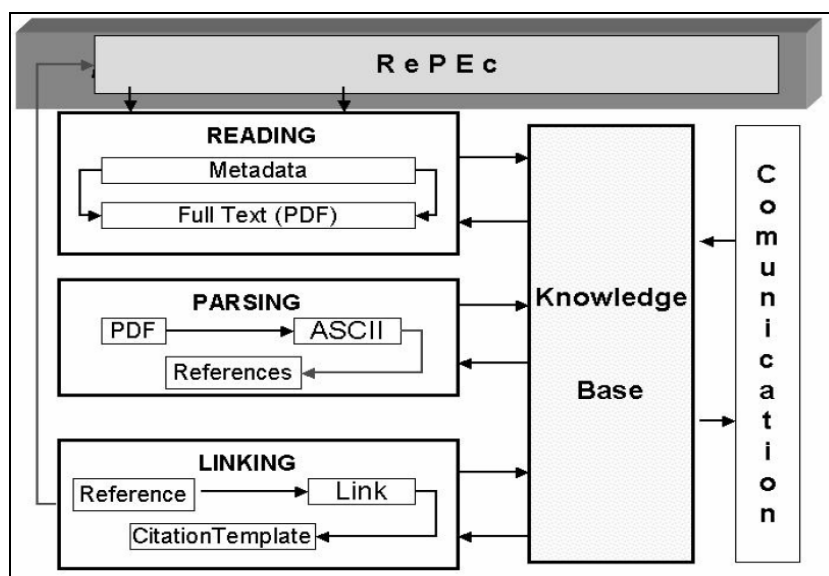
Dalším statistickým výstupem služby LogEc je počet nejvyhledávanějších autorů (kteří jsou zaregistrováni službou RePEc Author Service). Seznam je možné uspořádat podle stažených souborů nebo zobrazených abstraktů v období posledního měsíce, čtvrtletí, nebo celkové doby provozu RePEc. Dále služba LogEc umožňuje tento seznam filtrovat podle země původu autora.

4.2.4.8 CitEc

Služba **CitEc** (Citations in Economics, <http://citec.repec.org/>) je zatím nejmladší službou systému RePEc. Má experimentální povahu a je stále předmětem dalšího zdokonalování a vývoje.

Jejím cílem je stát se podstatným nástrojem pro realizaci citačních analýz dokumentů zahrnutých v databázi RePEc. Zpracování se týká zatím jen vybrané množiny dokumentů dostupných v elektronické podobě v rámci systému RePEc. Je potřebné zdůraznit, že systém RePEc nebudoval doposud centrální digitální fond úplných textů – množina úplných textů je však pro rozvoj služby CitEc podstatná. Jde zejména o pracovní studie (working papers). Podmínkou automatického zpracování citací z dostupných textů a následně citačního rejstříku nad digitálním fondem je, že dokumenty jsou ve formátu PDF nebo PostScript (PS) – naštěstí oba formáty jsou v oblasti ekonomiky populární. Služba CitEc totiž využívá v praxi již zavedenou a osvědčenou technologii **Autonomního citačního indexování** ACI (Autonomous Citation Indexing), která byla vyvinuta tvůrci slavné digitální knihovny CiteSeer (dnes CiteSeerX). Podrobnosti zpracování citací a citačního indexu v systému CiteSeer (ResearchIndex) jsou popsány v konferenčním článku autorky [BRATKOVÁ, 2001].

Služba CitEc je provozována na serveru Technické univerzity ve Valencii (Španělsko) a je řízena především jedním ze zakladatelů systému RePEc J.M. Barruecem, pracovníkem Univerzitní knihovny této univerzity. Projekt služby byl popsán také v článku v časopise „The Grey Journal“ v roce 2005 [BARRUECO, 2005].



Obr. 4/17: schéma zpracování citací a referencí subsystému CitEc systému RePEc (převzato z [BARRUECO, 2005])

Specifikem a jistou výhodou (a zároveň rozdílem od postupu v systému CiteSeer) zpracování citací z digitálních dokumentů v subsystému CitEc je existence rozsáhlé bibliografické databáze RePEc s relativně kvalitními, ale co do počtu údajů různorodými metadaty o zdrojových dokumentech (viz část „Knowledge Base“ ve schématu na obr. 4/17). Kvalita metadat je jedním z problémových momentů citačního zpracovatelského procesu [BARRUECO, 2005, s. 93], což je způsobeno i tím, že formát ReDIF zadává jen velmi málo povinných údajů, nepovinným údajem je například datum zveřejnění, který tak citelně chybí při zpracování citací. Vážným problémem distribuovaného systému jsou mnohdy neplatná URL. Metadata jsou využívána pro lepší identifikaci a sjednocování forem citací

extrahovaných z úplných textů dokumentů za pomoci technologie ACI. Programátoři systému CitEc navíc připravili 3 doplňkové programové moduly pro potřeby zpracování citací a zejména propojování záznamů (referencí):

- Modul pro sběr metadat a zejména úplných textů dokumentů (z lokálních archivů systému RePEc, zatím jen o výběr dokumentů, jejich získávání pro potřeby zpracování na centrálním serveru je mnohdy problém [BARRUECO, 2005, s. 95]) – viz blok „Reading“ ve schématu na obr. 4/17; elektronické texty zdrojových dokumentů musejí být konvertovány do ASCII kódu (pomocí programu pstotext) pro potřeby další lexikální analýzy a extrakce citací a referencí
- Modul pro textovou analýzu dokumentů (blok „Parsing“ ve schématu) za účelem extrakce citací uváděných v textu a referencí lokalizovaných na konci textů a identifikaci jejich jednotlivých údajů s využitím metadatové databáze RePEc
- Modul propojování citací (referencí) (blok „Linking“ ve schématu) s elektronickými dokumenty, pokud jsou zařazeny v pracovním digitálním fond systému, který dostal označení „**RePEC2**“

Ke konci roku 2008 služba CitEc zahrnovala celkem **180 480** zpracovaných zdrojových dokumentů, **3 979 807** metadatových záznamů a **1 716 086** citací.

Uživatelé systému RePEc a jeho služby CitEc mohou již využívat prvních výsledků budovaného subsystému.

Data služby CitEc jsou ale také poskytována dalším službám systému RePEc s tím, že tyto služby mohou zdokonalit své služby pro uživatele a badatelskou obec. Následující služby systému RePEc již zařadily do svých služeb citační data: Socionet, EconPapers a IDEAS.

4.2.5 Závěr

Digitální knihovna distribuovaného typu RePEc je nepochybně jedním z nejzajímavějších systémů moderní éry elektronické komunikace informací v oblasti vědy a výzkum se zaměřením na ekonomickou oblast. Jeho čelním představitelům se daří udržovat dobrý stav celého systému a jeho četných služeb i zajišťovat jeho další rozvoj. Je to dáno nadšením a až „heroickým“ úsilím řady talentovaných ekonomů a programátorů z různých částí světa (USA, Velké Británie, Švédsko, Španělsko, Rusko aj.), kteří pracují ve virtuálním prostoru. (Minulost informačních služeb zaznamenala řadu heroických snah jednotlivců, které nakonec skončily buď zánikem, nebo předáním informačního produktu profesionálním komerčním firmám). Celý systém pak „zaměstnává“ cca dalších 100 dobrovolníků, kteří zajišťují procesy a operace v lokálních účastnických archivech. Systém však potřebuje finanční zajištění, a to se děje formou sponzorství řady institucí, které působí v oblasti.

Pro informační profesionály či knihovníky, kteří tradičně a i dnes zajišťují pro vědu a výzkum organizování a zpřístupňování informací, vyvstává v konfrontaci se systémem RePEc řada otázek. Nastupuje snad trend, že si sami odborníci (vědci, badatelé) vystačí v budoucnu sami při organizování informací, že budou zajišťovat komplexní informační služby? Může zaniknout specializace informačního profesionála? Vydrží heroické úsilí budovatelů RePEc do budoucna? Změní se architektura systému, například ve prospěch budování centralizovaného digitálního fondu, který je potřebný pro budování citačních služeb? Kdo zajistí trvalou ochranu digitálních informací včetně digitálních textů do budoucna? Přežijí klasické komerční bibliografické registrace typu EconLit do budoucna? Odpovědi na otázky by bylo možné hypoteticky sdělit, ovšem vývoj nás může v mnoha směrech překvapit.

4.3 Elektronické archivy a digitální knihovny oboru knihovní a informační vědy

Mezi velkým počtem archivů elektronických tisků (lokální, národní nebo mezinárodní úrovně) s otevřeným přístupem lze v současné době identifikovat již několik archivů zaměřených na obor knihovní a informační vědy. Jsou určeny k rychlé a bezplatné komunikaci informací z tohoto oboru. Jejich otevřenost nespočívá jen v rovině **bezplatného vyhledávání a využívání** informací koncovými uživateli nebo dalšími informačními službami, ale také v rovině **svobodného a trvalého ukládání** vlastních dokumentů a jiných informací samotnými autory. Archivy elektronických tisků patří v současnosti, vedle volně dostupných online časopisů, k základním dvěma formám otevřeného elektronického publikování ve vědě.

4.3.1 Archiv elektronických tisků E-LIS

V oboru knihovní a informační vědy patří dnes k nejvýznamnějším a největším archivům mezinárodní centralizovaný archiv *Elektronických tisků v knihovní a informační vědě* (Eprints in Library and Information Science, E-LIS, <http://eprints.rclis.org/>). Pro české autory je zajímavý, kromě jiného, především proto, že na rozdíl od dalších existujících archivů umožňuje ukládat dokumenty v **národních jazycích**, tedy také v **češtině**. Do současné doby se v rámci archivu vytvořily dílčí skupiny uživatelů podle jazykového zaměření. Doporučovaným jazykem, nicméně, zůstává angličtina. Povinností autorů je ale uložení abstraktu a klíčových slov také v **angličtině**. Archiv E-LIS má již v celosvětovém rámci renomé, jeho metadata jsou automaticky pomocí protokolu OAI-PMH přebírána řadou různých informačních systémů a služeb (service providers) včetně citačních, takže vyhledatelnost, využitelnost i citovanost uložených dokumentů se neustále zvyšuje.

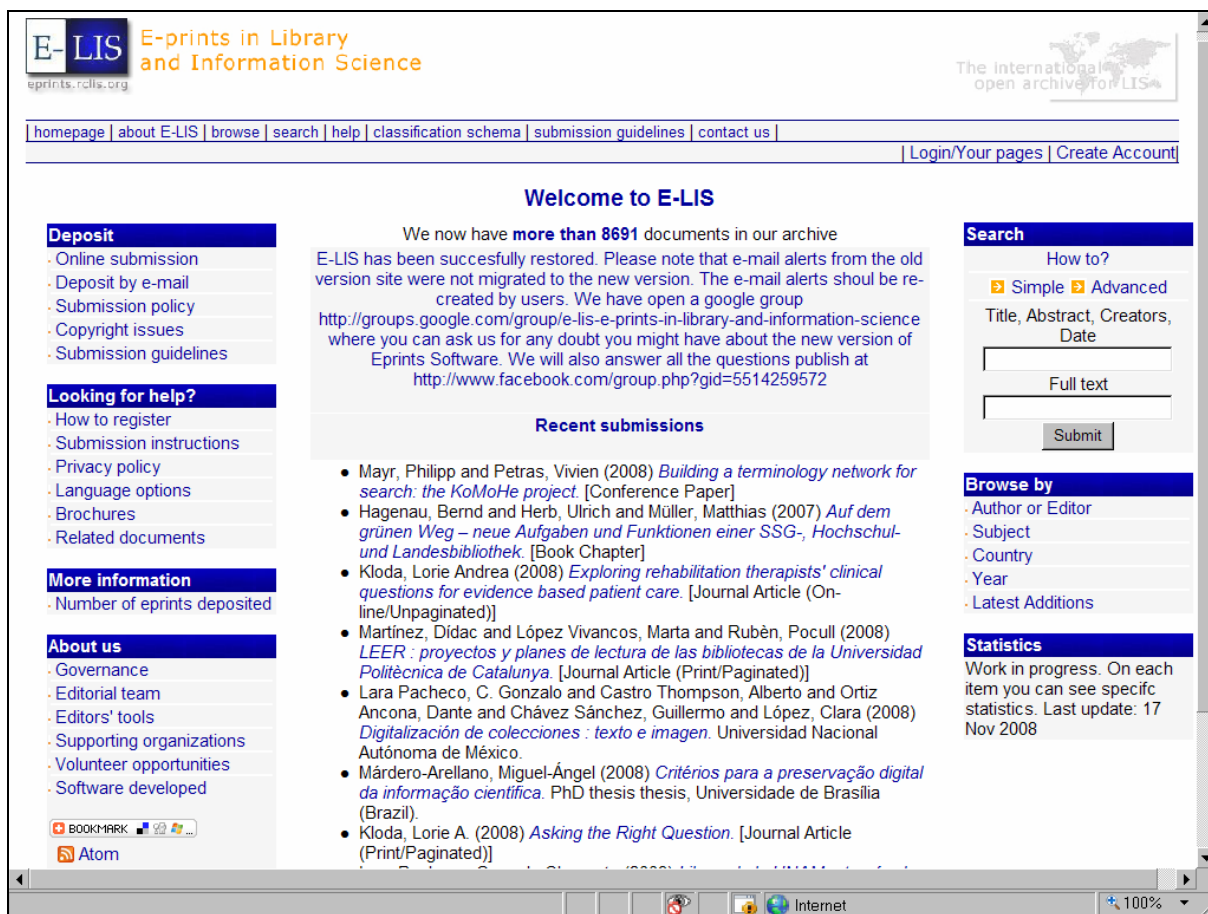
Archiv E-LIS byl do provozu uveden v lednu 2003. Jeho cílem bylo podpořit otevřený, bezplatný a snadný přístup k dokumentům reprezentujícím především výzkumné aktivity v oboru knihovní a informační vědy v globálním rámci. Archiv byl založen jako rozšíření služby **RCLIS** (Research in Computing, Library and Information Science, <http://rclis.org/>) a zejména jeho dílčí služby DoIS (<http://wotan.liu.edu/does/>). Má mezinárodní ráz a je určen neomezenému počtu autorů. Z hlediska obsahu prozatímni množiny uložených dokumentů by archiv E-LIS mohl být označen i jako **archiv evropský**, zhruba 60 % dokumentů pochází od autorů z Evropy, zejména ze Španělska, Itálie, Velké Británie a Německa. V systému narůstá i počet dokumentů psaných jiným typem písma než latinkou (například čínským písmem). Zahrnuta je publikovaná i nepublikovaná literatura z oboru knihovní a informační vědy včetně informačních technologií a příbuzných disciplín. Archiv má v současnosti **5 330** registrovaných uživatelů.

4.3.1.1 Organizační, programové a technické zabezpečení archivu

Archiv E-LIS má čistě nekomerční ráz a je virtuálně spravován a řízen skupinou dobrovolníků pocházejících z různých zemí. V jeho čele jsou **správci archivu** (I. Subirats Coll ze Španělska a A. De Robbio z Itálie). Aby bylo možné úspěšně zvládnout kontrolu narůstajícího počtu dokumentů z různých zemí a zejména dokumentů psaných v různých jazycích, bylo brzy po založení archivu přijato rozhodnutí o rozšíření počtu editorů. V principu jde alespoň o jednoho editora z každé zastoupené země, u zemí s velkým počtem autorů je stanoveno editorů více. Dnes působí v archivu okolo 60 editorů. Česká republika je v archivu zastoupena jedním editorem (jde o autorku této práce, která roli přijala i z důvodu bližšího obeznámení se s interním provozem celého systému, robor systému je tudíž detailnější). Hlavní úlohou editorů je kontrolovat ukládané dokumenty a metadata a schvalovat je k přesunu z pracovní

databáze (User Work Area) do „živého“ **archivu** (Live Archive). Metodické, technické, technologické a jiné otázky řeší editoři v rámci uzavřeného diskusního listu.

Archiv E-LIS nabízí uživatelům řadu funkcí (viz obr. č. 4/18). Koncovým uživatelům jsou k dispozici dílčí funkce pro **vyhledávání informací** typu **prohlížení záznamů pomocí rejstříků** (Browse) a typu **přímého dotazování** (Search) v režimu jednoduchém a pokročilém (Simple / Advanced) včetně vyhledávání v úplných textech. Autorům je po registraci (Create Account) umožněn přístup k trvalému ukládání dokumentů a tvorbě metadat (Login/Your pages). Systém zajišťuje i další **informační služby** (např. statistiky využívání uložených dokumentů, služby typu RSS, export záznamů do různých formátů). Archiv je rovněž připraven na sklizeň metadat na základě protokolu OAI-PMH. Od 20. 1. 2003 je zaregistrován v centrále iniciativy **OAI** (Open Archives Initiative) včetně bazového URL (<http://eprints.rclis.org/perl/oai2>). Metadata archivu jsou získatelná pro všechny návazné informační služby. Pro uživatele jsou připraveny **příručky** a pro náhodné zájemce je k dispozici již bohatá **dokumentace** včetně propagačních brožur a letáků. Specifické místo je věnováno prezentaci vlastní oborové klasifikace **JITA** (její popis viz dále v textu).



Obr. č. 4/18: domovská stránka elektronického archivu E-LIS s prezentací stručných záznamů nově uložených přírůstků [převzato 2008-12-19]

Archiv E-LIS je provozován v jednom z nejznámějších programů pro elektronické archivy - v programu britské provenience **EPrints.org** (<http://www.eprints.org/software/>). Od 21. listopadu 2008 je provozován v jeho **nejnovější verzi 3.1.0** (v termínu 17.-20. 11. 2008 proběhla komplexní migrace dat). Tento program je k dispozici zdarma pro nekomerční využití (GNU, General Public Licence). Patří k tzv. otevřeným zdrojům (Open Sources).

Program je možné upravovat a rozšiřovat. Vznikl na Fakultě elektroniky a počítačové vědy Southamptonské univerzity (Velká Británie) a jeho autorem je R. Tansley (verze 1). Rozvoj programu je finančně podporován britskou Společnou komisí pro informační systémy JISC (The Joint Information Systems Committee, <http://www.jisc.ac.uk/>) a americkou Národní vědeckou nadací UA (The National Science Foundation, National, <http://www.nsf.gov/>). Design a kódování 2. a 3. verze programu zajišťuje Christopher Gutteridge s dalšími spolupracovníky (Mike Jewell aj.).

Archiv je z hlediska technického zabezpečení v péči týmu Projektu pro pokročilá řešení elektronického publikování AePIC (<http://www.aepic.it/>) realizovaného v rámci Konsorcia italských univerzit CILEA (<http://www.cilea.it/>). Tato organizace je jednou z partnerských organizací archivu E-LIS a zajistila k jeho provozu počítačovou kapacitu. Pracovník týmu AePIC Z. Tajoli zajišťuje každodenní technickou správu a údržbu archivu. V rámci systému je využit operační systém Linux (RedHat 7.3), databáze je uložena v systému MySQL a prezentace informací na WWW je zajišťována dynamicky pomocí web serveru Apache a souboru skriptů v jazyce Perl. K dalším sponzorům či partnerům patří španělské Ministerstvo školství a vědy, Centrum pro výzkum digitálních knihoven na Strathclydské univerzitě (Velká Británie), britská informační služba BUBL, Brazílský institut pro informace ve vědě a technice (IBICT), Polská knihovnická asociace, Turecká knihovnická asociace aj.

4.3.1.2 Ukládání dokumentů do archivu E-LIS

4.3.1.2.1 Přijímané typy a formáty dokumentů

Archiv E-LIS přijímá relativně velké množství různých typů dokumentů z oboru knihovnictví a informační vědy a oborů souvisejících. Na rozdíl od některých dalších archivů není omezen jen na elektronické tisky, tj. **preprinty** nebo **postprinty vědeckých článků** určených v konečné fázi k publikování v recenzovaných komerčních nebo volně dostupných **časopisech**. Aktuální verze systému (Eprints 3.1.0) nabízí autorům celkem 23 typů dokumentů k ukládání do archivu:

- Preprint [Preprint, tj. nepublikovaný text]
- Thesis [Vysokoškolská kvalifikační práce libovolného typu]
- Book [Kniha, popř. jiný typ monografické práce]
- Book Chapter [Kapitola knihy, popř. jiné monografické práce]
- Bibliography [Bibliografie, v monografické formě]
- Guide/Manual [Průvodce/Příručka, v monografické formě]
- Tutorial [Výukový instrukční materiál, tištěný nebo elektronický]
- Library Instructional Material [Instrukční materiál určený pro práci v knihovně]
- Conference Proceedings [Konferenční sborník jako celek]
- Conference Paper [Stat' konferenčního sborníku]
- Conference Poster [Plakát určený k prezentaci informací na konferenci]
- Presentation [jen pro prezentaci bez textu konferenční statě]
- Project/Business Plan [Projektový/Obchodní plán]
- Report [Zpráva, jiného typu než níže uvedeně]
- Departmental Technical Report [Technické zprávy fakult nebo ústavů apod.]
- Technical Report [Technická nebo výzkumná zpráva]
- Journal Article (On-line/Unpaginated) [Časopisecký článek online nestránkovaný]

- Journal Article (Print/Paginated) [Časopisecký článek tištěný stránkovaný]
- Review [Recenze libovolného materiálu]
- Newspaper/Magazine Article [Článek z novin nebo populárního časopisu]
- In Collection [Skupina dokumentů tvořících edici nebo sbírku]
- Dataset [Datový soubor]
- Other (Jiný typ materiálu, který není zahrnut v základním výčtu)

Nabídka typů dokumentů je v nové verzi systému Eprints 3.1.0 systémem předkládána autorům na samém začátku popisu ukládaného dokumentu. Volba některé z kategorií vede v druhém kroku k zobrazení formuláře s různým počtem metadat podle povahy dokumentu. Nová verze zahrnuje oproti verzi předcházející obecný typ Zpráva (Report) a typ Recenze (Review), který dříve citelně chyběl. Z druhé strany není zahrnut typ Pracovní zpráva (Working Paper) a typ Syllabus (Syllabus).

Ukládané dokumenty mohou být co do publikačního stavu materiály již publikovanými, materiály, které mají být publikovány, nebo materiály nepublikovanými.

Zatímco u typů dokumentů v archivu E-LIS v zásadě neexistují omezení, u jejich formátů je tomu jinak. Dokumenty jsou přijímány s cílem jejich **kvalitního plnotextového vyhledávání** a zejména **trvalého uložení** (i nadále se v budoucnosti souvislosti se změnami technologií počítá s jejich další migrací, emulací apod.). Striktně doporučenými formáty jsou proto formáty **PDF** (včetně prezentací PPT) a **HTML** a popřípadě formát **TXT**. Třetí verze systému EPrints umožňuje již také seskupování různých digitálních objektů do jednoho celku (například text přednášky z konference, PPT prezentace a zvukového záznamu vystoupení). Takový celek je také připraven k exportování do metadatového formátu METS (obr. 4/23).

4.3.1.2.2 Aktuální stav repozitáře archivu

V archivu E-LIS se k polovině prosince 2008 nacházelo již téměř **8 700 dokumentů**. Archiv začal být naplňován v lednu 2003 a od té doby se počet přírůstku, především nově vzniklých dokumentů, pohybuje, v posledních 5 letech v průměru kolem **900 ročně** (viz hodnoty na obr. 4/19). Do archivu jsou ale ukládány i dokumenty vytvořené v dřívějším období a tak se dá hovořit také o jeho nepravidelném **retrospektivním naplňování**. Nejstarší uložené dokumenty jsou prozatím z roku 1952 a 1957.

1952 (1)	1983 (8)	1997 (156)
1957 (1)	1984 (4)	1998 (159)
1965 (1)	1985 (8)	1999 (264)
1966 (1)	1986 (11)	2000 (331)
1969 (1)	1987 (12)	2001 (459)
1974 (2)	1988 (12)	2002 (648)
1975 (1)	1989 (9)	2003 (902)
1976 (2)	1990 (13)	2004 (954)
1977 (1)	1991 (16)	2005 (1117)
1978 (5)	1992 (24)	2006 (1395)
1979 (1)	1993 (24)	2007 (1211)
1980 (1)	1994 (57)	2008 (656)
1981 (5)	1995 (102)	2009
1982 (4)	1996 (112)	2010

Obr. č. 4/19: přehled počtu dokumentů uložených v archivu E-LIS [k 2008-12-19]

Rejstřík zemí původu dokumentů, který je uživatelům k dispozici v rámci vyhledávání typu prohlížení (Browse by Country), vykazuje zároveň jednoduchou, ale zajímavou statistiku. Archiv zahrnuje dokumenty pocházející z různých zemí světa. Vzhledem k tomu, že je provozován v **Evropě**, je prozatím největší počet dokumentů pocházejících z tohoto kontinentu (5 019). Je potěšitelné, že archiv vzbudil v posledních letech velký zájem také u autorů ze **Severní a Střední Ameriky**, takže jejich dokumenty přibývají velkým tempem (prozatím více než 1 702 dokumentů). Velký počet dokumentů pochází již také z **Asie** (prozatím více než 954 dokumentů, aktivní jsou zejména čínští autoři) a z **Jižní Ameriky** (prozatím 958 dokumentů). Zhruba po 50-60 dokumentech je zastoupeno ze zemí **Afriky** a **Austrálie a Oceánie**.

Z evropských zemí zauímají přední místa zejména dokumenty autorů ze **Španělska** (více než 1666 dokumentů), **Itálie** (více než 984), **Velké Británie** (více než 358), **Turecka** (cca 287) a **Německa** (více než 260). Následují další země s menším množstvím uložených dokumentů, Česká republika je zastoupena 18 jednotkami.

4.3.1.3 Metadata pro zpracování ukládaných dokumentů

V rámci procesu ukládání dokumentu do archivu má autor k dispozici pohodlné komplexní rozhraní pro správu ukládání dokumentů (Manage Deposits). Autor je povinen vyplnit pomocí pohodlného webového formuláře **metadata** popisující dokument. Autoři jsou v pokynech i náповědách vyzýváni, pokud možno, k preciznímu uvádění dat z důvodu kvalitního online vyhledávání. Novinkou nové 3. verze systému je využívání autorizovaných dat z externích databází (například pro jména autorů), pokud existují. Každý typ dokumentu má jinou skladbu metadat. K základním formálním popisným metadatovým prvkům patří:

- Publikáční stav (Publication status) [Unpublished, In Press, Published]
- Recenzovaný/Nerecenzovaný dokument (Referend, Yes/No)
- Veřejná doména (Public domain, Yes/No)
- Autoři (Authors)
- Název dokumentu (Title)
- Editoři [sborníků apod.] (Editors)
- Nakladatel nebo vydavatel (Publisher)
- Rok [vytvoření nebo publikování knihy nebo časopisu] (Year)
- Země [původu dokumentu] (Country)
- Jazyk dokumentu (Language)
- Edice (Series)
- Konference [úplné jméno] (Conference)
- Datum konference (Conference date)
- Místo konference (Conference location)
- Alternativní lokace zdroje [URL včetně protokolu] (Alternative locations)
- Doplnkové informace [typu poznámkové věty] (Additional information)
- Komentář k [dokumentu uloženému v archivu] (Commentary on)
- Bibliografické odkazy [záznamy citované literatury] (References)

U analytických dokumentů se vyplňují:

- Kapitola [číslo kapitoly v knize] (Chapter)
- Publikace [zdrojový dokument knižní nebo časopisecký] (Publication)
- Recenzovaný časopis (Refereed)
- Ročník [u časopisu] (Volume (journal))
- Číslo [u časopisu] (Number)
- Strany [rozsah] (Pages)

U specifických typů dokumentů (šedé literatury) se zpravidla uvádí:

- Typ vysokoškolské práce (Thesis type)
- Instituce (Institution)
- Oddělení (Department)

K metadatům **věcného charakteru** patří:

- Abstrakt [v národním jazyce a angličtině] (Abstract))
- Předmětové termíny [slovní hodnoty klasifikace JITA] (Subjects)
- Klíčová slova [v národním jazyce a angličtině] (Keywords)

Některá **formální metadata** se uvádějí v národním jazyce podle originálu (název aj.), některá musejí být v angličtině (místo konání konference, jméno instituce aj.), hodnoty některých údajů se přebírají z číselníků (jsou v angličtině). Klasifikace JITA (viz další podkapitola) má slovní hodnoty vyjádřené také v angličtině.

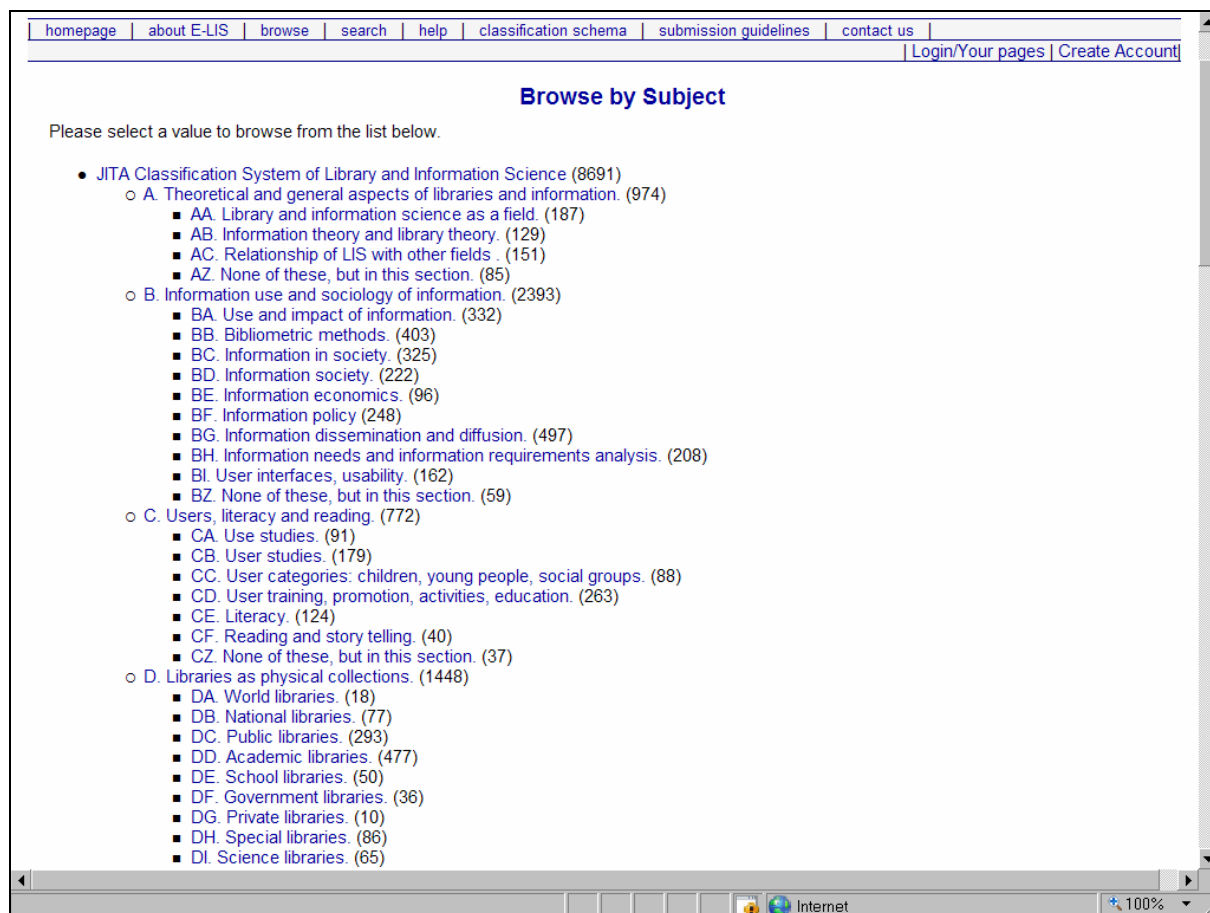
Před vyplňováním metadat je vhodné připravit si předem **seznam použité (citované) literatury**, protože se vkládá do jednoho pole formuláře (v samotném dokumentu zůstává také). Struktura záznamů citované literatury je na autorovi (prozatím není předepsán jednotný styl). Platí ale pravidlo, že citovaný záznam by měl mít všechny základní údaje, protože archiv E-LIS využívá externích služeb k propojování citací buď na úplné texty nebo bibliografické záznamy v jiných zdrojích. Systém prozatím není připraven na automatickou extrakci citací z textů.

4.3.1.4 Oborová klasifikace JITA

Základním systémem pro věcné pořádání dokumentů i pro jejich online vyhledávání je v archivu E-LIS jeho **klasifikace JITA** (<http://eprints.rclis.org/jita/>) pro obor knihovní a informační vědy. Byla připravena na základě sloučení původní klasifikace „*NewsAgentTopic*“, kterou vytvořil a do konce března 1998 udržoval Mike Keen (Aberystwyth, Velká Británie), a také klasifikace již nefunkčního německého časopisu *Review of Information Science (RIS)*, kterou původně vytvořil Američan Donald Soergel. Zkratka **JITA** pochází z křestních jmen zakladatelů archivu E-LIS (**J**osé Manuel Barrueco Cruz, **I**mma Subirats Coll, **T**homas Krichel a **A**ntonella De Robbio).

Klasifikace JITA zahrnuje v současné době 12 hlavních předmětových kategorií, označených písmeny velké anglické abecedy **A-L** (část klasifikace viz na obr. č. 4/20) a doplněných slovní hodnotou. Kategorie jsou v celkovém schématu zařazeny do tří úrovní (ty nejsou ale prezentovány v rámci předmětového rejstříku). Kategorie **A** a **B** patří do **obecné úrovně** (zahrnují teoretické a obecné otázky oboru a otázky využívání a sociologie informací), kategorie **C, D, E, F, G** do **střední úrovně** (zahrnují otázky uživatelů informací, typů knihoven jako fyzických fondů, otázky publikování a jeho právních aspektů, otázky managementu informací a lidských zdrojů, informačního průmyslu, profese a výchovy v oboru) a kategorie **H, I, J, K, L** do **úrovně specifické** (zahrnují otázky informačních zdrojů, jejich nosičů a kanálů, otázky zpracování informací pro informační služby, otázky technických služeb v knihovnách, archivech a muzeích a otázky informačních a knihovnických technologií). Základní kategorie jsou na druhé úrovni děleny do podkategorií, kterých je v současné době celkem 135 (v počátečním období jich bylo 124). (Za dva poslední roky nepřibyla žádná nová podkategorie [BRATKOVÁ, 2006a, s. 6]). Počet podkategorií v rámci základních kategorií se různí (mezi 4 až 20) a jsou označovány také pomocí písmen abecedy (např. HB - šedá literatura aj.). Klasifikace má dynamickou povahu, vždy poslední podkategorie je označena písmenem Z (např. DZ=Žádná z uvedených [podkategorií], ale [bude] v této sekci). Autor může tuto poslední kategorii v rámci tvorby

metadat zvolit, pokud nenachází potřebnou podkategorii ve stanoveném schématu. Správci klasifikace průběžně situaci monitorují a vyhodnocují a v případě většího počtu zpracovaných dokumentů s takovou podkategorií rozhodují o vytvoření podkategorie nové a o jejím zařazení do schématu.



Obr. č. 4/20: část schématu klasifikace JITA prezentované v rámci předmětového rejstříku archivu E-LIS [převzato 2008-12-19]

O charakteru klasifikace JITA je možné diskutovat, nicméně je třeba ocenit úsilí tvůrců při pokusu o nové oborové pořádání informací. Je dokladem toho, že ani nové systémy se bez profesionálního pořádání nemohou obejít, ačkoliv i na to nejsou jednotné názory.

Po téměř šestiletém provozu archiv E-LIS obsahuje k polovině prosince 2008 již téměř 8 700 dokumentů (viz obr. 4/20). Aktuální přehled předmětových kategorií prozrazuje, že největší množství uložených dokumentů náleží především do tříd specifické úrovně. Největší obsazení má kategorie „H“ Informační zdroje, nosiče a kanály (téměř 2 474 dokumentů), kategorie „B“ Využívání informací a sociologie informací (více než 2 390), kategorie „L“ Informační a knihovnické technologie (1 670), kategorie „I“ Zpracování informací pro informační služby (1 508), a kategorie „D“ Knihovny jako fyzické fondy (téměř 1 448).

4.3.1.5 Ukládání informací

Proces ukládání dokumentů do archivu je vázán na **registraci autora** (Create Account), která je zdarma stejně jako využívání veškerých poskytovaných služeb. Registrace se provádí ve dvou krocích. V prvním kroku je nutné vyplnit vstupní registrační formulář v rozhraní „Register“ (<http://eprints.rclis.org/cgi/register>). Podstatné je vyplnění identifikačního jména

a hesla. Ve druhém kroku autor registrační proces dokončí tak, že po obdržení e-mailové zprávy a jejím potvrzení vyplní v systému po zalogování se v rozhraní „Login“ (<http://eprints.rclis.org/cgi/users/home>) svůj **uživatelský záznam**. Důležitá je také přesná e-mailová adresa pro komunikaci editorů archivu s autory. Svůj uživatelský záznam může autor kdykoliv upravovat. Některé informace mají veřejnou povahu, jsou spojeny s ukládanými dokumenty a budou využívány také v online vyhledávání. Před zahájením tohoto procesu, stejně jako před vlastním ukládáním dokumentu, je žádoucí, aby se autor seznámil s příslušnými příručkami, které jsou k dispozici na webových stránkách archivu.

★ Authors

Please enter the authors' names or select from the list. Is your name missing in this list? You should register to the International Registry for Authors in Library and Information Science (IraLIS). The aim of IraLIS is to create a register of authorities for the authors in Library and Information Science. Please be registered at <http://www.iralis.org>

	Family Name	Given Name / Initials	Person ID	
1.	Bratková	Eva	CZLIS0685	<input checked="" type="checkbox"/>
2.	Mach	Jan		<input type="checkbox"/>
3.				<input type="checkbox"/>

[More input rows](#)

★ Title

Please enter the full title of the deposit.

Standardizace metadat pro národní registr EVŠKP

★ Subjects

Please select at least one main subject category, and optionally up to two other subject categories you think are appropriate for your submission, in the list below.

[Remove](#) HP. e-resources.

[Remove](#) IE. Data and metadata structures.

Search for subject: [Search](#) [Clear](#)

- [Add](#) ☒ A. Theoretical and general aspects of libraries and information.
- [Add](#) ☒ B. Information use and sociology of information.
- [Add](#) ☒ C. Users, literacy and reading.
- [Add](#) ☒ D. Libraries as physical collections.
- [Add](#) ☒ E. Publishing and legal issues.
- [Add](#) ☒ F. Management.
- [Add](#) ☒ G. Industry, profession and education.
- [Add](#) ☒ H. Information sources, supports, channels.
- [Add](#) ☒ I. Information treatment for information services.

Obr. č. 4/21: část vstupního webového formuláře pro ukládání popisných metadat (verze 3.1.0) v archivu E-LIS s vyplněnými údaji o dokumentu českých autorů [převzato z WWW 2008-12-18]

Ukládání informací do archivu musí být zahájeno v rozhraní pro **správu ukládání dokumentů** (Manage deposits). Autor může proces ukládání kdykoliv přerušit, aniž by se data ztratila. V rámci pracovního prostoru autor neustále vidí seznam záznamů, které jsou rozpracovány, a seznam již dokončených záznamů, které čekají na přesun do hlavního archivu po schválení editorem. Tyto záznamy autor může editovat, vymazávat, klonovat, exportovat do jiných formátů nebo může vytvářet záznamy nové.

Vlastní ukládací proces zahrnuje několik kroků, reprezentovaných dílčími částmi webového formuláře. V prvním kroku se v rámci tvorby nového záznamu vybírá z předdefinované nabídky **typ** ukládaného dokumentu (23 typů). Volba typu dokumentu ovlivňuje skladbu metadat, která se objeví v dalších částech formuláře. V druhém kroku se vyplňuje základní sada **popisných metadat** o ukládaném dokumentu, a to buď přímým zápisem (užitečné je kopírování dat z dokumentu), nebo označením zvoleného údaje v nabídkovém seznamu, jako

je seznam **klasifikace JITA** (v poli „Subject“), jazyk dokumentu (pole „Language“), země dokumentu (pole „Country“) aj. Novinkou verze 3.1.0 systému EPrints je možnost přebírání autorských údajů včetně trvalého identifikátoru z databáze mezinárodního registru autorů z oblasti knihovni a informační vědy **IraLIS** (<http://www.iralis.org>), která je spravována španělskými odborníky. Autoři jsou vyzýváni k registraci v tomto systému dříve než začnou do systému vkládat metadata a dokumenty. Ukázka části vyplněných metadat popisujících práci českých autorů je na obr. 4/21.

Zvláštní pozornost musí být věnována poli **abstrakt** (Abstract). Povinností je vyplnit abstrakt v angličtině. U neanglicky psaných dokumentů je požadován také abstrakt v jazyce dokumentu, který se v nové verzi konečně ukládá do zvláštního opakovatelného pole (Other abstract) – dříve existovalo jen pole jediné. Je možné také připojovat abstrakta v dalších jazycích (jazyk se vybírá z rolovací nabídky číselníku archivu E-LIS).

V dalším kroku je autor požádán o podání (upload) počítačového souboru buď přímo z počítače autora nebo zápisem URL, na kterém se soubor nachází (mělo by jít vždy o soubor vytvořený autorem, nikoliv nakladatelem). Závěrečné potvrzení funkce „**Finished**“ (Ukončeno) vede k zobrazení náhledu celého záznamu v prezentačním formátu. Autor by měl metadata pečlivě překontrolovat a v případě nalezené chyby se vrátit k provedení opravy. Definitivní potvrzení uložení dokumentu se realizuje pomocí funkce „**Deposit**“. Systém v nové třetí verzi umožňuje v rámci jednoho záznamu nahrávat k jednomu záznamu různé typy objektů (například text, video a zvuk). Hlídá také případné duplicitní ukládání dokumentů. Možný je také import metadat z jiných systémů.

4.3.1.6 Autorská práva

Všechny práce uložené v archivu zůstávají majetkem autora (majetková práva). Pokud autor uloží do volně dostupného archivu **nerecenzovaný preprint** vědeckého článku, drží všechna práva k jeho rozmnožování a rozšiřování (copyright). Autor ručí za to, že ukládaný dokument není spojen s žádnými omezeními stran jeho elektronického rozšiřování. K auto-archivaci autor nepotřebuje povolení od nikoho ani v případě **předběžně recenzovaného preprintu**. Jde-li o **recenzovaný postprint**, pak se autor může pokusit modifikovat dohodu o předání copyrightu podepsanou s nakladatelem s cílem umožnit uložení postprintu ve volně dostupném, archivu. Není-li dohoda s vydavatelem možná, může autor k uloženému preprintu připojit opravy ve zvláštním souboru.

Právo auto-archivace recenzovaného postprintu je právní záležitostí, neboť dohoda o předání copyrightu se týká textu tohoto postprintu. Ale předběžně recenzovaný preprint je ukládán do archivu v době, kdy dohoda o předání copyrightu ještě neexistuje, a to znamená, že je to autor, kdo drží exkluzivně a zcela copyrightová práva. Mnoho nakladatelů autorům auto-archivaci povoluje, ale záleží na politice daného nakladatele. Autor také může do archivu uložit postprint, k němuž povolí omezený přístup. Je případně na autorovi, aby požádal nakladatele o návrat určité části copyrightových práv, například uložit postprint do volně dostupného archivu nebo na vlastní webový server. Někteří vydavatelé proklamují, že tato práva standardně poskytují (Elsevier aj.). Copyrightová práva dávají tvůrcům exkluzivní práva, která mohou být jak rozdělována na dílčí části, tak předávána jiným osobám.

Specifická situace se týká metadat ukládaných do archivu. Celosvětově je v současné době umožněno bezplatné **sklizení metadat** (nikoliv dokumentů samotných) z archivů elektronických tisků za účelem budování nadstavbových systémů či služeb, jejichž úkolem je, kromě jiného, podpora vyhledávání a prezentace obsahů těchto archivů.

Dokumenty uložené v archivu jsou chráněny na základě copyrightu, který drží autor. Autoři kontrolují práva případného překladu a reprodukce svých prací. Nicméně **koncoví uživatelé**

mají v rámci otevřeného přístupu dovoleno využívat těchto děl pro osobní, výchovné nebo výzkumné účely. Mohou také tato díla rozmnožovat na nekomerční bázi. U uživatelů se předpokládá **férové užití** děl z archivů, stejně jako tomu je u děl publikovaných tradiční cestou. Dokumenty z archivů smějí být čteny v režimu online, přebírány na počítačová média pro osobní využití, URL lokalizace těchto dokumentů smí být hyperlinkována (citována) v jiných dokumentech. Dokument ale nesmí být bez souhlasu autora měněn.

4.3.1.7 Výstupy a služby archivu E-LIS

Archiv E-LIS v nové verzi 3.1.0 systému EPrints nabízí škálu **služeb a výstupů** pro koncové i jiné uživatele. K základním službám patří **online vyhledávání informací**. Je dostupné ve dvou základních režimech: prohlížení záznamů pomocí rejstříků vybraných selekčních údajů a dotazovací vyhledávání pomocí formulace rešeršního dotazu. Vyhledávání informací je anonymní, uživatelé se nemusejí registrovat.

4.3.1.7.1 Online vyhledávání a zobrazování informací

Prohlížení záznamů pomocí rejstříků

Metoda prohlížení záznamů pomocí rejstříků vybraných selekčních údajů je vhodná v případě, kdy uživatel neví přesně, co chce hledat. Jde o intuitivní způsob vyhledávání. Nabídku rejstříků lze získat přímo na hlavní stránce (vpravo) nebo ze specifického rozhraní „prohlízet“ (browse, <http://eprints.rclis.org/view/>). K dispozici jsou v rámci nové verze 3.1.0 již jen **čtyři rejstříky**: (rejstřík názvů konferencí a rejstřík názvů knih a časopisů z předcházející verze byly zrušeny):

- Jména autorů a editorů (Author or Editor)
- Předmětové termíny (Subject) (Jde o klasifikaci JITA)
- Jména zemí (Country)
- Roky vytvoření nebo vydání dokumentu (Year)


V rámci prohlížení rejstříků lze zjišťovat i jednoduché **statistiky**. U jednotlivých selekčních údajů jsou v závorkách uváděny počty jim odpovídajících záznamů (viz například počet záznamů v rámci předmětových kategorií klasifikace JITA na obr. č. 4/20). Z položek selekčních údajů se uživatel dostává na druhé nebo dalších úrovních nejdříve k seznamu zkrácených záznamů a poté k **rozšířenému prezentačnímu záznamu** uloženého dokumentu (viz obr. č. 4/22). Po získání tohoto záznamu lze přejít buď k **úplnému textu** (uloženému v repozitáři archivu nebo popřípadě i na jiném externím úložišti – „Alternative Locations“), nebo navigovat dále pomocí hyperodkazů na další záznamy či seznamy (podle předmětové kategorie). U řady záznamů lze využít funkce propojení na záznamy nebo přímo dokumenty, které byly autorem daného dokumentu citovány (přes odkaz **SEEK**, který se nachází na konci každé citace). Pomocí hyperlinku „**Show all fields**“ lze v nové verzi získat výpis úplného metadatového záznamu (z databáze SQL) v tabulkové formě (Full Metadata – jeho vizualizace však zahrnuje některé potíže, například se nezobrazují URL u citací). Zajímavou vlastností vizualizace informací je zobrazování statistiky přístupu k záznamu a dokumentu ([View statistics for this eprint](#)) – viz obr. č. 4/22.

Specifickým rejstříkem je seznam záznamů **nových přírůstků za poslední týden** (Recent submissions). Podobný ale časově širší seznam se skrývá pod hyperodkazem „**Latest Additions**“ (napravo v hlavní stránce systému).

Knowledge Technologies

Milton, Nick (2008) *Knowledge Technologies*. Publishing Studies, 3. POLIMETRICA.

Full text available as:

 PDF - Requires a PDF viewer such as GSview, Xpdf or Adobe Acrobat Reader
903Kb Language: English

Abstract(s)

Several technologies are emerging that provide new ways to capture, store, present and use knowledge. This book provides a comprehensive introduction to five of the most important of these technologies: Knowledge Engineering, Knowledge Based Engineering, Knowledge Webs, Ontologies and Semantic Webs. For each of these, answers are given to a number of key questions: What is it? What can it be used for? What tools are available? What are the main issues? What are the main practitioners interested in Knowledge Management, Artificial Intelligence, Design Engineering and Knowledge Engineering? The book is written for students, researchers and practitioners at the University of Nottingham on the application of AI techniques to knowledge management and the development of expert systems for military applications. In 1999, he joined Epistemics where he worked to establish knowledge management programmes at large organisations in the engineering, technology and business sectors. "Knowledge Acquisition in Practice", which describes a step-by-step procedure for acquiring and developing expert systems with leading research organisations working on knowledge technologies, such as knowledge-based engineering, Knowledge Engineering, Knowledge Based Engineering, Knowledge Webs, Ontologies and Semantic Webs.

Item Type: Book

Keywords: Knowledge Engineering, Knowledge Based Engineering, Knowledge Webs, Ontologies and Semantic Webs

Subjects: L. Information technology and library technology.
H. Information sources, supports, channels.
A. Theoretical and general aspects of libraries and information.

Full Metadata: [Show all fields](#)


ID Code: 12800

Deposited By: Sica, Viviana

Deposited On: 27 Feb 2008

Last Modified: 19 Nov 2008 12:58

Statistics: [View statistics for this eprint](#)



Preview

http://eprints.rclis.org/12800/1/Knowledge_Technologies_reading.pdf

Obr. č. 4/22: výstupní zkrácený metadatový záznam knižní publikace uložené v archivu E-LIS s odkazem na statistiku přístupu [převzato 2008-12-18]

Vyhledávání pomocí formulace dotazů

Přímé formulování dotazu lze realizovat několika způsoby. Prvním je (rychlé) **vyhledávání** (Search) přímo z hlavní stránky, a to podle slov z pole názvu, abstraktu, jmen autorů a data publikování. Připojit lze i termíny k prohledávání indexu úplných textů dokumentů (Full text).

Druhou možností je volba specifického rozhraní pro **jednoduché vyhledávání** (Simple Search, <http://eprints.rclis.org/cgi/search/simple>), ve které je možné realizovat vyhledávání podle slov z názvu dokumentu, abstraktu, jmen autorů a roku zveřejnění dokumentu opět s případným připojením termínů pro fulltextové vyhledávání. Tento způsob má navíc k dispozici nabídku třídění získaných záznamů. Třetím je **pokročilé vyhledávání** (Advanced Search, <http://eprints.rclis.org/cgi/search/advanced>), které nabízí velkou škálu selekčních údajů: vyhledávání v úplných textech dokumentů, dále podle slov z názvu dokumentu, slov jmen autorů i editorů, slov z abstraktu, klíčových slov, předmětových kategorií, typu dokumentu, země původu dokumentu, jazyka, publikačního stavu, recenzního stavu (Refereed), názvu zdrojového dokumentu, roku zveřejnění).

Formulace dotazů se provádí pomocí formulářů. Některé selekční údaje lze zapisovat přímo, jiné se vybírají z nabídkových seznamů. Využít lze jednoduchých **booleovských operací**: „All of“ (AND) nebo „Any of“ (OR).

Novinkou třetí verze systému EPrints je možnost exportu nalezených metadatových záznamů do různých metadatových formátů (Dublin Core, MODS, METS), do citačních manažerů (EndNote, BibTeX aj.) nebo do jiných systémů a jejich databází. V archivu E-LIS je v nové

verzi je také možné kontejnerové sdružování souvisejících objektů - na obr. č. 4/23 je záznam konferenční statě autorky a jejího spoluautora (viz [BRATKOVÁ, 2008b]) ve strukturovaném formátu METS; zahrnuje bibliografický popis titulu ve formátu MODS v části „mets:dmdSec“, informaci k autorským právům v části „mets:amdSec“, popis jednotlivých souvisejících objektů (text a 2 prezentací) v části „mets:fileSec“ a souhrn všech souborů v části „mets:structMap“.

```

<mets:mets OBJID="oai:eprints.rclis.org:15633"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:mods="http://www.loc.gov/mods/v3" LABEL="Eprints Item"
  xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-0.xsd"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:mets="http://www.loc.gov/METS/">
  <mets:metsHdr CREATEDATA="2009-03-28T11:50:33Z">
    <mets:agent TYPE="ORGANIZATION" ROLE="CUSTODIAN">
      <mets:name>E-LIS</mets:name>
    </mets:agent>
  </mets:metsHdr>
  <mets:dmdSec ID="DMD_oai:eprints.rclis.org:15633_mods">
    <mets:mdWrap MDTYPE="mods">
      <mets:xmlData>
        <mods:titleInfo>
          <mods:title>Metadatový standard EVSKP-MS, v. 1.1 pro popis
            VŠKP a standardy související</mods:title>
        </mods:titleInfo>
        <mods:name type="personal">
          <mods:namePart type="given">Eva</mods:namePart>
          <mods:namePart type="family">Bratková</mods:namePart>
        <mods:role>
          <mods:roleTerm type="text">author</mods:roleTerm>
        </mods:role>
        </mods:name>
        <mods:name type="personal">
          <mods:namePart type="given">Jan</mods:namePart>
          <mods:namePart type="family">Mach</mods:namePart>
        <mods:role>
          <mods:roleTerm type="text">author</mods:roleTerm>
        </mods:role>
        </mods:name>
        <mods:abstract>The article is devoted to presentation of national
          Electronic Theses and Dissertations Metadata Set in the Czech
          Republic (EVSKP-MS, version 1.1) and related standards for
          Persons description (PersCZ, version 1.0) and Corporate bodies
          description (CorpCZ, verze 1.0) too. All standards were prepared
          within the Centralized project „National registry of ETDs and ETD
          repository with plagiarism detecting service“ (MŠMT C1/2008). In
          introduction part purposes, significance and starting points of all
          metadata sets creation are characterized. Both, the second and
          third parts contain brief description of descriptive, technical and
          administrative metadata elements specifications, mandatory
          elements for purposes of their exchange on national and
          international level are emphasized. Last part explains the XML
          schema of metadata records and metadata transfer among
          systems using OAI-PMH protocol.</mods:abstract>
        <mods:classification authority="lcc">IE. Data and metadata
          structures.</mods:classification>
        <mods:classification authority="lcc">HB. Gray
          literature.</mods:classification>
        <mods:classification authority="lcc">HP. e-
          resources.</mods:classification>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:dmdSec>
  <mets:originInfo>
    <mods:dateIssued encoding="iso8061">2008-10-
      07</mods:dateIssued>
  </mets:originInfo>
  <mets:originInfo>

```

```

    <mods:publisher>Odborná komise pro otázky elektronického
    zpřístupňování VŠKP AKVŠ ČR</mods:publisher>
  </mods:originInfo>
  <mods:genre>Conference Paper</mods:genre>
</mets:xmlData>
</mets:mdWrap>
</mets:dmdSec>
<_ <mets:amdSec ID="TMD_oai:eprints.rclis.org:15633">
  <_ <mets:rightsMD ID="rights_oai:eprints.rclis.org:15633_mods">
    <_ <mets:mdWrap MDTYPE="mods">
      <_ <mets:xmlData>
        <_ <mods:useAndReproduction>
          <_ <p>
            <strong>For work being deposited by its own
            author:</strong>
            In self-archiving this collection of files and associated
            bibliographic metadata, I grant E-LIS the right to
            store them and to make them permanently available
            publicly for free on-line. I declare that this material is
            my own intellectual property and I understand that
            E-LIS does not assume any responsibility if there is
            any breach of copyright in distributing these files or
            metadata. (All authors are urged to prominently
            assert their copyright on the title page of their work.)
          </p>
          <_ <p>
            <strong>For work being deposited by someone other
            than its author:</strong>
            I hereby declare that the collection of files and
            associated bibliographic metadata that I am archiving
            at E-LIS) is in the public domain. If this is not the
            case, I accept full responsibility for any breach of
            copyright that distributing these files or metadata
            may entail.
          </p>
          <p>Clicking on the deposit button indicates your
            agreement to these terms.</p>
        </mods:useAndReproduction>
      </mets:xmlData>
    </mets:mdWrap>
  </mets:rightsMD>
</mets:amdSec>
<_ <mets:fileSec>
  <_ <mets:fileGrp USE="reference">
    <_ <mets:file SIZE="388031" ID="oai:eprints.rclis.org:15633_11094_1"
      MIMETYPE="application/octet-stream"
      OWNERID="http://eprints.rclis.org/15633/1/ETDmetadataCZ.pdf">
      <mets:FLocat LOCTYPE="URL"
        xlink:href="http://eprints.rclis.org/15633/1/ETDmetadataCZ.pdf"
        xlink:type="simple" />
    </mets:file>
  </mets:fileGrp>
  <_ <mets:fileGrp USE="reference">
    <_ <mets:file SIZE="409717" ID="oai:eprints.rclis.org:15633_11095_1"
      MIMETYPE="application/octet-stream"
      OWNERID="http://eprints.rclis.org/15633/2/Bratkova1.pdf">
      <mets:FLocat LOCTYPE="URL"
        xlink:href="http://eprints.rclis.org/15633/2/Bratkova1.pdf"
        xlink:type="simple" />
    </mets:file>
  </mets:fileGrp>
  <_ <mets:fileGrp USE="reference">
    <_ <mets:file SIZE="167078" ID="oai:eprints.rclis.org:15633_11096_1"
      MIMETYPE="application/octet-stream"
      OWNERID="http://eprints.rclis.org/15633/3/Mach1.pdf">
      <mets:FLocat LOCTYPE="URL"
        xlink:href="http://eprints.rclis.org/15633/3/Mach1.pdf"
        xlink:type="simple" />
    </mets:file>
  </mets:fileGrp>
</mets:fileSec>

```

```

</mets:fileGrp>
</mets:fileSec>
<mets:structMap>
  <mets:div DMDID="DMD_oai:eprints.rclis.org:15633_mods"
    AMDID="TMD_oai:eprints.rclis.org:15633">
    <mets:fptr FILEID="oai:eprints.rclis.org:15633_11094_1" />
    <mets:fptr FILEID="oai:eprints.rclis.org:15633_11095_1" />
    <mets:fptr FILEID="oai:eprints.rclis.org:15633_11096_1" />
  </mets:div>
</mets:structMap>
</mets:mets>

```

Obr. 4/23: metadatový záznam článku z konference z archivu E-LIS v exportovacím strukturovaném formátu METS [převzato 2008-12-13]

Obr. 4/24: rozhraní pro nastavení služby Alerts v archivu E-LIS [převzato 2008-12-22]

4.3.1.7.2 Další služby archivu

Registrovaní uživatelé mohou využívat **signální služby** (typu Alerts) pomocí elektronické pošty na základě zadaných profilů. Je nutné nejdříve uložit formulaci dotazu pomocí funkce „Saved Searches“ a nastavit poté ve formuláři několik parametrů, jako je název rešerše, frekvence zasílání informací aj. – viz obr. 4/24.

Pro neregistrované uživatele se nabízejí také služby typu RSS (RDF Site Summary) v několika verzích:

RSS 1.0 (http://eprints.rclis.org/cgi/latest_tool?output=RSS)

RSS 2.0 (http://eprints.rclis.org/cgi/latest_tool?output=RSS2)

Atom (http://eprints.rclis.org/cgi/latest_tool?output=Atom)

Přihlásit se k odběru novinek archivu lze již z hlavní stránky archivu. Zabezpečuje dodávání metadat o nových přírůstcích do archivu.

Metadatové záznamy z archivu E-LIS jsou v současnosti sklízeny řadou jiných informačních služeb (včetně internetových vyhledávačů, jako je Gogole a jeho služby Google Scholar). Důležitá je například externí služba **METALIS** (<http://metalisis.cilea.it/>), která umožňuje vyhledávání nad mnoha archivy z oboru knihovnictví a informační vědy. Metadata se dostávají také do důležitého britského systému **Citebase Search** (<http://www.citebase.org/>), který kromě vyhledávání záznamů a dokumentů umožňuje sledování citovanosti dokumentů.

Výše zmíněná služba **propojování citovaných dokumentů** s jejich úplnými texty je novější službou archivu. Služba je realizována na základě specifického programu **ParaCite Toolkit** (zkráceně **ParaTools**, version 1.00, <http://paracite.eprints.org>), který je v současné chvíli integrální součástí programu Eprints. Program rozkládá záznamy citací na jednotlivá metadata a směřuje je ke službě s webovým formulářem pro vyhledávání příslušných dokumentů ve vybraných volně dostupných zdrojích na WWW. Patří k nim databáze již zmíněného systému CiteBase Search (<http://www.citebase.org/>), databáze systému CiteSeerX (<http://citeseerx.ist.psu.edu/>), databáze indexu vyhledávače Google (<http://www.google.com>) jeho citační službě Google Scholar (<http://scholar.google.com/>). K propojení s úplnými texty se využívá standardu **OpenURL**. Do hry mohou v budoucnu vstoupit zdroje další.

4.3.1.8 Závěr

Archiv E-LIS je v současné chvíli jedním z nejvýznamnějších mezinárodních archivů s otevřeným přístupem v oboru knihovnictví a informační vědy. Poskytuje řadu zajímavých výstupů a služeb. Externí citační služby zajišťují větší využívání uložených informací. Archiv zahrnuje již relativně velký počet různorodých typů dokumentů v elektronické formě včetně metadat. Jeho velkým specifickým je velká variabilita zastoupených jazyků, ve kterých jsou dokumenty napsány. Jde o jeden z mála archivů, do kterých mohou autoři uložit dokument ve svém mateřském jazyce, tedy také češtině. Archiv je zcela připraven na sklizení metadat jinými službami.

4.3.2 Archiv elektronických tisků pro vědy o informacích a komunikaci „@rchiveSIC“

Francouzský archiv elektronických tisků nazvaný Archiv pro vědy o informacích a komunikaci **@rchiveSIC** (Archive en Sciences de l'Information et de la Communication, SIC, <http://archivesic.ccsd.cnrs.fr/>) byl založen v jarních měsících roku 2002. Do úplného provozu byl uveden jako jeden z prvních v našem oboru již v červnu 2002. Jedním z podnětů k založení tohoto archivu byly alarmující výsledky vzešlé z ankety provedené v té době francouzským ministerstvem výzkumu. Anketa sledovala využívání vědeckých či odborných periodik v oblasti věd o informacích a komunikaci (SIC). Bylo zjištěno, že ze 263 světových periodik pouze 17 titulů bylo známo alespoň jednomu odborníkovi ze dvou. Z 40 titulů přímo francouzských periodik jich bylo známo jenom 12 alespoň jednomu ze dvou odborníků [GALLEZOT, 2003, Des revues SIC].

Odpovědí na tuto situaci byl návrh na založení otevřeného archivu pro vědy o informacích a komunikaci. Jeho tvůrci a správci zároveň, tj. Gabriel Gallezot z Národní vysoké školy věd o informacích a knihovnách (ENSSIB) v Lyonu, Ghislaine Chartronová z Regionální jednotky pro výchovu v oblasti vědeckých a technických informací (URFIST) v Paříži a Jean-Max Noyer z Univerzity Paříž 7, využili příznivé situace vzniku Centra pro přímou vědeckou

komunikaci (Centre pour la Communication Scientifique Directe, <http://ccsd.cnrs.fr>) v rámci francouzského Národního centra pro vědecký výzkum (CNRS), které již v té době provozovalo archiv pro elektronické disertace (<http://tel.archives-ouvertes.fr/>) a zrcadlové sídlo světového archivu pro přírodní vědy arXiv.org (<http://fr.arxiv.org/>). Archiv @rchiveSIC byl založen na serveru výše zmíněného Centra [GALLEZOT, 2002, část VI]. Na jeho založení se dále významně podílely jednotky URFIST v Paříži a v Nice, Výzkumná skupina pro informační služby (GRESI), která do roku 2003 pracovala v rámci ENSSIB, a Výzkumné centrum pro komunikaci obrazových informací Oddělení pro vzdělávání v oblasti kinematografie, komunikace a informací (CCI/CRECI) na Univerzitě Paříž 7.

Francouzský archiv byl jedním z prvních, který v oboru informační věda adaptoval software **EPrints**. Byl provozován v jeho verzi 1.1.1. Ve srovnání s jinými archivy měl však tehdy však lepší design prezentace informací a výstupů. Jako jediný tehdy také umožňoval navíc vyhledávání v úplných textech. Adaptaci softwaru pro potřeby @rchiveSIC připravil Christian Rossi. Archiv byl také 19. 12. 2002 zaregistrován v centrále iniciativy OAI (<http://www.openarchives.org/Register/BrowseSites.pl>). Na této URL adrese je však dnes již modifikovaná informace. Od **11. 8. 2006** byl obsah archivu migrován **do nového softwarové aplikace francouzské provenience HAL** (Hyper Article en Ligne, <http://hal.archives-ouvertes.fr/>). Nová báze adres tohoto archivu pro sklizeň metadat je již změněná: (<http://archivesic.ccsd.cnrs.fr/oai/oai.php>). Základní identifikace archivu je na URL: (<http://archivesic.ccsd.cnrs.fr/oai/oai.php?verb=Identify>). Metadata archivu jsou dosažitelná pro různé návazné informační služby včetně vyhledávacích služeb, jako je Google.

Archiv @rchiveSIC je primárně určen komunitě frankofonních učitelů a badatelů z oblasti věd o informacích a komunikaci, lze ho tedy označit za mezinárodní (občas bývá označován jako archiv národní povahy). Dominuje v něm v současné chvíli francouzština, ukládat dokumenty je ale možné také v jazycích dalších (anglickém, německém, italském nebo španělském) [GALLEZOT, 2003]. K uložení elektronického dokumentu je nutná, tak jako v dalších archivech, registrace autora, který poskytuje základní údaje o sobě a na základě nich obdrží ze strany systému ID a heslo pro vstup do modulu pro ukládání dat. V rámci ukládacího procesu autor vyplňuje metadata o dokumentu pomocí webového formuláře.

Seznam **předmětových kategorií archivu @rchiveSIC** (viz jejich seznam na obr. 4/25) odráží aktuální přístup francouzských odborníků k vymezení oboru, který u nás označujeme termínem „informační věda“. Na rozdíl od amerického pojetí (prezentovaného například i v rámci archivu DLIST), které rozšiřuje název oboru v současné době o oblast informačních technologií (Information Science and Technologies), je francouzské pojetí typické jednak tím, že jde o více věd o informacích, jednak tím, že je název rozšířen v tomto případě o slovo „komunikace“ (se zdůrazněním na komunikaci ve vědě). V předmětové kategorizaci @rchiveSIC se tak objevuje řada příbuzných oborů (muzeologie, studia masmédií, sociologie, ekonomika, umění, estetika aj.).

[Bibliométrie, scientométrie \(35\)](#)

[Cinéma, art, esthétique \(58\)](#)

[Collectivités territoriales \(75\)](#)

[Communication et information scientifique \(180\)](#)

[Conflits, Stratégie, Veille \(40\)](#)

[Documentation \(141\)](#)

[Droit de l'information/communication \(19\)](#)

[Economie, industries culturelles \(55\)](#)

[Edition électronique \(140\)](#)

[Education, formation \(85\)](#)

[Espace public \(63\)](#)
[Géopolitique \(20\)](#)
[Gestion des connaissances \(83\)](#)
[Histoire de l'information/communication \(44\)](#)
[Hypertextes, hypermédia \(99\)](#)
[Ingénierie des systèmes d'information \(163\)](#)
[Médias de masse \(84\)](#)
[Muzeologie \(21\)](#)
[Organisation et communication \(128\)](#)
[Sociologie de l'information/communication \(176\)](#)
[Théories information/communication \(168\)](#)
[Autres \(13\)](#)

Obr. 4/25: předmětové kategorie archivu @rchiveSIC s počty dokumentů [stav k 2008-12-19]

Obr. 4/25 odráží aktuální stav počtu uložených dokumentů i jejich rozdělení do předmětových kategorií z hlediska obsahu. K 15. prosinci 2008 bylo v archivu uloženo celkem cca **991** dokumentů, jimž bylo dohromady přiděleno 630 předmětových kategorií (v průměru je jeden dokument opatřen dvěma kategoriemi). Nejvíce dokumentů je zařazeno pod kategorii Komunikace a vědecké informace (180), dále Sociologie informací a komunikace (176 dokumentů zahrnujících zejména výsledky ze sociologických výzkumů), dále pod kategorií Teorie informací a komunikace (168), Inženýrství informačních systémů (163) a Elektronické publikování (140).

Věcný popis dále zahrnuje **abstrakt** a **volně tvořená klíčová slova** ve francouzštině a popřípadě také v angličtině.

Archiv @rchiveSIC je sice primárně určen pro frankofonní odborníky, rozhraní systému je ale připraveno i v jazyce anglickém (včetně seznamu předmětových kategorií) pro uživatele z jiných jazykových oblastí. Poskytuje, kromě komplexních služeb pro autory ukládající dokumenty, řadu dalších **standardních i nadstavbových služeb**. Jde především o **vyhledávání informací**. Systém hned na úvodní stránce nabízí krátký seznam **aktuálně uložených dokumentů**, pokud byly uloženy. V rámci vyhledávání v režimu prohlížení nabízí zajímavý **rejstřík 10 nejvyužívanějších dokumentů** k danému datu a samozřejmě také rejstříky dokumentů řazených podle **autorů a předmětových kategorií**. V tomto archivu však, na rozdíl od ostatních, chybí rejstřík dokumentů upořádaných podle roku vytvoření/zveřejnění. **Přímé zadávání dotazu** je možné jak v jednoduchém režimu (podle slov z názvu a abstraktu ve francouzštině nebo klíčových slov ve francouzštině, názvu časopisu nebo konferenčního sborníku, autorů a předmětových kategorií dle výběru z řízeného seznamu), tak v režimu pokročilém (podle názvu dokumentu, autorů, názvu časopisu nebo konferenčního sborníku, slov z abstraktu ve francouzštině a angličtině a klíčových slov, předmětových kategorií dle výběru z řízeného seznamu a typu dokumentu). Specifikem francouzského archivu je vyhledávání uložených dokumentů podle identifikátoru (celého identifikátoru nebo jenom čísla). Je vhodné k rychlému nalezení záznamu a dokumentu, ale předpokládá se již dobrá orientace v archivu ze strany uživatele. Zajišťuje také **vyhledávání v úplném textu dokumentu** (v rámci aplikace softwaru HAL).

Zobrazované prezentační záznamy uložených dokumentů obsahují na konci jedinečné identifikátory uložených dokumentů (včetně jejich varianty „oai“) a včetně informace o verzi textu. Na základě identifikátoru „oai“ je možné získat metadatový záznam pomocí protokolu OAI-PMH přes libovolnou službu, například známý systém „Repository Explorer“ (RE) (<http://re.cs.uct.ac.za/>). Záznam zobrazený na obr. 4/26 reprezentuje dokument typu zpráva (v první verzi). Má přidělený systémový identifikátor: **sic_00345575_v1**

[Accueil](#)
[Mon espace](#)
[Consulter](#)
[Rechercher](#)
[Déposer](#)
[Services](#)
[?](#)

Fiche détaillée

Liste des fichiers attachés à ce document :

PDF
rapport-decembre-2008-final-Chartron-Caillon-1.pdf (843.2 KB)

sic_00345575, version 1

Citations des ressources électroniques dans les publications scientifiques : Analyse comparée

Ghislaine Chartron¹, Elisabeth Caillon²
(10/2008)

Ce rapport s'intéresse aux nouvelles dimensions de la citation dans les textes scientifiques dans le contexte numérique. A l'appui d'une étude exploratoire quantitative menée fin 2007 et début 2008 dans le domaine de la biologie moléculaire, de la physique appliquée et de l'économie-gestion, les résultats montrent une faible intégration de ressources Web dans les textes (en dehors des versions en ligne des vecteurs traditionnels), sauf pour la biologie moléculaire. L'étude met en évidence une typologie des mécanismes de liens développés par les éditeurs et les agrégateurs. Un ensemble de recommandations finales sont formulées pour accompagner cette économie du lien devenue aujourd'hui centrale pour les citations scientifiques.

1 : INTD-CNAM - Institut national des techniques de la documentation
 2 : INTD-CNAM - Institut National des techniques de la documentation

Domaines scientifiques : **Sciences de l'Homme et Société/Sciences de l'information et de la communication/Communication et information scientifique**

Mots-clés en français : **citation – référence bibliographique – publication scientifique – ressource numérique – éditeur – agrégateur – biologie moléculaire – physique appliquée – sciences sociales – sciences économiques – économie des liens**

sic_00345575, version 1
http://archivesic.ccsd.cnrs.fr/sic_00345575/fr/
[oai:archivesic.ccsd.cnrs.fr:sic_00345575_v1](http://archivesic.ccsd.cnrs.fr/sic_00345575_v1)
 Mardi 9 Décembre 2008, 13:19:40

Obr. 4/26: výstupní prezentační záznam zprávy uložené v bázi archivu @rchiveSIC včetně identifikátorů na spodním okraji obrazovky [převzato 2008-12-22]

Jeho varianta pro potřeby komunikace přes protokol OAI-PMH je:

oai:archivesic.ccsd.cnrs.fr:sic_00345575_v1

Na obr. 4/27 je zobrazen výpis metadatového záznamu dokumentu s výše uvedeným identifikátorem ve standardním formátu Dublin Core (oai_dc), získaný přes službu RE (ve struktuře XML). Záznam obsahuje datum uložení dokumentu do archivu - v hlavičce záznamu (2008-12-09T13:19:40Z).

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2009-03-26T15:38:11Z</responseDate>
  <request verb="GetRecord" metadataPrefix="oai_dc"
    identifier="oai:archivesic.ccsd.cnrs.fr:sic_00345575_v1">http://archivesic.ccsd.c
    nrs.fr/oai/oai.php</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:archivesic.ccsd.cnrs.fr:sic_00345575_v1</identifier>
        <datestamp>2008-12-09T13:19:40Z</datestamp>
        <setSpec>SHS:INFO:COMM</setSpec>
      </header>
    </record>
  </GetRecord>
</OAI-PMH>
```

```

<= <oai_dc:dc
  xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:identifier>HAL:sic_00345575, version 1</dc:identifier>
    <dc:identifier>http://archivesic.ccsd.cnrs.fr/sic_00345575/en/</dc:identifier>
  <dc:title>Citations des ressources électroniques dans les publications scientifiques : Analyse comparée</dc:title>
  <dc:creator>Chartron, Ghislaine</dc:creator>
  <dc:creator>Caillon, Elisabeth</dc:creator>
  <dc:subject>Humanities and Social Sciences/Library and information sciences/Scientific communication and information</dc:subject>
  <dc:subject>citation</dc:subject>
  <dc:subject>référence bibliographique</dc:subject>
  <dc:subject>publication scientifique</dc:subject>
  <dc:subject>ressource numérique</dc:subject>
  <dc:subject>éditeur</dc:subject>
  <dc:subject>agrégateur</dc:subject>
  <dc:subject>biologie moléculaire</dc:subject>
  <dc:subject>physique appliquée</dc:subject>
  <dc:subject>sciences sociales</dc:subject>
  <dc:subject>sciences économiques</dc:subject>
  <dc:subject>économie des liens</dc:subject>
  <dc:description>Ce rapport s'intéresse aux nouvelles dimensions de la citation dans les textes scientifiques dans le contexte numérique. A l'appui d'une étude exploratoire quantitative menée fin 2007 et début 2008 dans le domaine de la biologie moléculaire, de la physique appliquée et de l'économie-gestion, les résultats montrent une faible intégration de ressources Web dans les textes (en dehors des versions en ligne des vecteurs traditionnels), sauf pour la biologie moléculaire. L'étude met en évidence une typologie des mécanismes de liens développés par les éditeurs et les agrégateurs. Un ensemble de recommandations finales sont formulées pour accompagner cette économie du lien devenue aujourd'hui centrale pour les citations scientifiques.</dc:description>
  <dc:publisher>HAL - CCSD</dc:publisher>
  <dc:date>2008</dc:date>
  <dc:type>research report</dc:type>
  <dc:language>FR</dc:language>
    <dc:relation>http://archivesic.ccsd.cnrs.fr/docs/00/34/55/75/PDF/rapport-decembre-2008-final-Chartron-Caillon-1.pdf</dc:relation>
  <dc:rights />
</oai_dc:dc>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

```

Obr. 4/27: metadatový záznam z archivu @rchiveSIC ve struktuře XML exportovaného dle protokolu OAI-PMH pomocí příkazu GetRecord (oai_dc) [převzato 2009-03-25]

Pro registrované uživatele nabízí archiv **diskusní fórum**. Uživatelé mohou také zasílat komentáře k uloženým dokumentům. Standardní službou je i poskytování **signálních služeb** (typu Alert) na základě zadaných profilů uživatelů pomocí elektronické pošty s tím, že uživatel sice získává seznam metadatových záznamů, ale může si bezprostředně v prostředí WWW okamžitě stáhnout úplný text. Služba **poskytování metadatových záznamů** dalším vyhledávacím a jiným informačním službám na základě protokolu OAI-PMH již byla zmíněna výše v textu (viz exportovaný záznam na obr. 4/27).

Nezbytnou službou profesionálního charakteru je poskytování **statistik** uložených dokumentů a uživatelů archivu, kteří buď ukládají dokumenty, nebo pouze odebírají signální výstupy, a také využívání (či návštěvnosti) archivu, která má stoupající tendenci. Tvůrci a správci archivu se již nyní zamýšlejí nad otázkou, jak docílit většího nárůstu dokumentů - navrženo je například ukládání všech statí z kolokvií a konferencí a paralelní ukládání článků z odborných časopisů [GALLEZOT, 2003].

4.3.3 Archiv elektronických tisků pro informační vědu a technologie DLIST

Třetím archivem elektronických tisků v oboru informační věda je americká Digitální knihovna informační vědy a technologií **DLIST** (Digital Library of Information Science and Technology, <http://dlist.sir.arizona.edu>), která byla uvedena do úplného provozu v říjnu 2002 (testování probíhalo již od 20. června 2002) ve spolupráci Školy informačních zdrojů a knihovní vědy a Arizonského centra pro zdravotnické vědy na Arizonské univerzitě (USA). Lze si všimnout, že v tomto případě je ve jménu archivu elektronických tisků uvedeno slovo „digitální knihovna“. Souvisí to zřejmě jak s rozšířeným záběrem přijímaných typů digitálních dokumentů, tak s budoucím plánovaným zaměřením celého systému vůbec [COLEMAN, 2003c]. K hlavním tvůrcům a správcům archivu patřila původně Anita S. Colemanová a Paul Bracke z výše uvedených institucí. Základním cílem uvedených institucí bylo vytvoření otevřeného, přes WWW dostupného digitálního repozitáře pro uvedený obor. I když se ve jménu archivu vyskytuje pouze termín „informační věda“ (vedle termínu „technologie“), jeho tvůrci často v dokumentaci a literatuře zdůrazňují širší termín „knihovní a informační věda“ (Library and Information Science, LIS) [COLEMAN, 2003a].

Dnes je odpovědnou redaktorkou archivu Cheryl Malonová (Arizonská univerzita), která patřila do původního týmu zakladatelů archivu. Kvalitu metadat, zejména z hlediska věcného popisu zajišťuje skupina dalších 12 odborníků z různých amerických univerzit.

Systém elektronického archivu je provozován v tomto případě zatím jen ve verzi 2.2 programové aplikace **EPrints.org**, který je připraven na sklizeň metadat ze strany návazných služeb na základě protokolu OAI-PMH (bázové URL: (<http://dlist.sir.arizona.edu/perl/oai2>)). Systém běží pod operačním systémem Linux. Databáze je uložena v systému MySQL a prezentace informací na WWW je zajišťována pomocí webserveru Apache a programovacího jazyka Perl. Doplnkový program Webalizer (otevřený zdroj) zajišťuje analýzy a zprávy o využívání archivu ze strany uživatelů.

Typologie zahrnovaných materiálů (dokumentů) je poměrně velmi široká (archiv nabízí v současné době vyhledávání pomocí rejstříku typů dokumentů). Zahrnuty jsou různé typy digitálních materiálů, jeho specifikem je větší množství typů materiálů určených pro potřeby **výuky a informační gramotnosti**. Do archivu jsou zařazovány bibliografické instruktážní materiály (bibliografie a přehledové seznamy), knihy a jejich kapitoly, edice monografických materiálů, konferenční sborníky, konferenční referáty (344 ke konci roku 2008), datové soubory (například z bibliometrických výzkumů), zprávy fakult nebo ústavů ze sledovaných oborů, průvodce, interaktivní materiály, časopisy (online stránkované i nestránkované) a jejich články (cca 390 ke konci roku 2008), knihovnické instruktážní materiály, články z novin a magazínů, preprinty, prezentace (cca 160 ke konci roku 2008), technické a výzkumné zprávy (cca 60), disertace (cca 20), sylaby, osnovy kursů a plány cvičebních lekcí, výukové materiály (k počítačovým programům a k tvorbě a využívání databází) a jiné. Pokud jde o publikační status ukládaných materiálů, může jít o materiály již publikované (nutné je dodání dokumentu v digitální formě vytvořené autorem, nikoliv vydavatelem, např. u souborů PDF), materiály, které mají být publikovány, nebo i materiály nepublikované.

V archivu DLIST se k polovině prosince 2008 nacházelo cca **1 377 dokumentů**. Archiv začal být naplňován systematicky v lednu 2003 a od té doby se počet přírůstku, především nově vzniklých dokumentů, pohybuje v průměru jen kolem **100-150 ročně**. (v posledních pěti letech kromě roku 2008 – viz obr. č. 4/28). Do archivu jsou pozvolna ukládány také dokumenty vytvořené v dřívějším období (nepravidelné **retrospektivní naplňování** – obr. č. 4/28). Nejstarší uložené dokumenty jsou z roku 1931 a 1951.

1931 (1)	1987 (3)	1998 (37)
1951 (2)	1988 (7)	1999 (42)
1961 (1)	1989 (4)	2000 (63)
1963 (1)	1990 (12)	2001 (64)
1966 (1)	1991 (6)	2002 (95)
1967 (1)	1992 (20)	2003 (107)
1973 (2)	1993 (14)	2004 (103)
1977 (1)	1994 (17)	2005 (175)
1979 (1)	1995 (20)	2006 (298)
1981 (2)	1996 (41)	2007 (153)
1985 (1)	1997 (19)	2008 (63)

Obr. č. 4/28: přehled počtu dokumentů uložených v archivu DLIST [k 2008-12-12]

Archiv DLIST má velmi malý nárůst dokumentů ve srovnání s jinými archivy oboru. Co do velikosti se zařazuje na druhé místo za archiv E-LIS, nicméně jeho velikost nenesrovnatelně menší. Příčin, proč je americký archiv tak málo v pozornosti většího počtu autorů z oblasti informační vědy, může být více. Jednou z nich může být i fakt, že obor je v USA zabezpečen velkým počtem komerčních produktů (včetně časopisů a sborníků) a služeb, které jsou jim snadno dostupné. Autoři tudíž nejsou, alespoň prozatím, motivováni k ukládání kopií dokumentů, pokud by tomu i vydavatel nebránil, do volně dostupného archivu.

Archiv DLIST dnes zahrnuje dokumenty v angličtině [COLEMAN, 2003a, část 1]. Autor ukládá elektronický dokument přímou cestou na server (uploading), popřípadě může specifikovat URL, z něhož může být dokument přenesen pracovníky archivu. V obou případech bude kopie díla uložena v databázi systému DLIST. Autor zároveň vyplňuje metadata, jejichž současná specifikace vychází z Dublin Core. Zahrnuje (prvky uvedené kurzivou nejsou povinné):

1. Typ elektronického tisku (viz položky typologie materiálů)
2. *Novější verze k (identifikátor předchozí verze dokumentu uloženého v archivu)*
3. *Komentář k (identifikátor komentovaného dokumentu uloženého v archivu)*
4. Status (nepublikován, v tisku, publikován)
5. *Veřejná doména (vyplňuje se, jde-li o starší a až nyní zveřejňovaný dokument jiného autora)*
6. *Jména editorů*
7. Název (dokumentu uloženého v archivu)
8. Předmětové kategorie (povinně jedna kategorie, maximálně dvě navíc volitelně)
9. Rok (zveřejnění nebo výroby ukládaného dokumentu)
10. *Abstrakt*
11. Konference (úplné jméno konference)
12. *Edice (název edice nebo souboru knih)*
13. *Datum konference*
14. *Lokace (místo konání konference)*
15. *Ročník (časopisu, v němž byl zveřejněn článek)*

16. Číslo (časopisu, v němž byl zveřejněn článek)
17. Nakladatel
18. Komentář k (úplný bibliogr. záznam komentovaného dokumentu uloženého v jiném archivu)
19. Alternativní lokace (URL jiných míst ukládaného dokumentu)
20. Klíčová slova (volně tvořená)
21. Doplnkové informace
22. Bibliografické odkazy (seznam citované/odkazované literatury určený pro potřeby tvorby citačního rejstříku)
23. Poznámky a návrhy (určené editorům archivu)
24. Formát (PDF, HTM, HTML, XLS, PPT, TEXT, Rich Text Format, PostScript, ASCII, DOC)
25. Doplnkové informace k formátu uloženého dokumentu
26. Stupeň bezpečnosti (lze limitovat přístup k dokumentu pouze pro personál archivu)

Po vyplnění metadat připojuje autor elektronický soubor, popř. více souborů, a ukončuje celou vstupní proceduru odesláním příslušné funkce směrem do systému. Autor tím také předává práva na uložení svého dokumentu v archivu a na jeho volné využívání ze strany uživatelů. Je však nezbytné, aby si autor sám hlídal případnou kolizi práv ke zveřejnění dokumentu, zejména pokud jde o komerční vydavatele jeho dokumentu.

Academic Libraries (76) Anthropology (4) Archaeology (3) Archives (80) Artificial Intelligence (45) Bibliometrics (58) Book History (58) Cataloging (31) Citation Analysis (46) Classification (112) Co-citation Analysis (6) Cognitive Science (22) Communications (14) Community Informatics (11) Computational Linguistics (2) Computer Science (15) Data Mining (21) Database Searching Instructions (12) Databases (28) Digital Humanities (13) Digital Libraries (186) Digital Preservation (8) Distributed Learning (30) Economics (4) Economics of Information (23) Electronic Publishing (48) Epistemology (14) Evaluation (60) Geographic Digital Libraries (4) Geographic Information Science (6) Geography (1) Government Information (18)	History (67) Human Computer Interaction (74) Hypertext and Hypermedia (46) Indexing (32) Information Analysis (17) Information Architecture (16) Information Ethics (22) Information Extraction (42) Information Literacy (42) Information Retrieval (39) Information Science (91) Information Seeking Behaviors (70) Information Systems (76) Informetrics (23) Interdisciplinarity (21) Internet (72) Journalism (1) Knowledge Management (48) Knowledge Organization (99) Knowledge Representation (17) Knowledge Structures (24) Learning Science (15) Libraries (68) Library and Information Science Education (141) Library Instruction (16) Library Science (50) Library Statistics (6) Library Systems (27) Linguistics (3) Management (16) Management Information Systems (8)	Map Librarianship (2) Media Studies (2) Medical Libraries (21) Metadata (28) Museums (22) Natural Language Processing (7) Neuroscience Ontology (4) Philosophy (21) Psychology (16) Public Libraries (11) Qualitative Research (15) Quantitative Research (11) Reading (2) Reference Services (43) Research Methods (12) Scholarly Communication (134) Science Technology Studies (57) Social Epistemology (13) Social Informatics (78) Sociology (27) Software (22) Standards (8) Training (15) User Studies (56) Virtual Communities (15) Web Metrics (15) Web Mining (19) Wireless Technologies (1) World Wide Web (80) XML (5) Z39.50
---	---	--

Obr. 4/29: Taxonomie archivu DLIST v rámci rejstříku vyhledávacího systému [převzato 2008-12-12]

Obsah repozitáře archivu DLIST tvoří v tuto chvíli digitální dokumenty z oblasti knihovnictví a informační vědy (Library and Information Science, LIS) a informačních technologií (Information Technologies, IT). Věcný popis zahrnuje především zařizování dokumentu do alespoň jedné z nabízených předmětových kategorií. **Taxonomie** archivu DLIST, prezentovaná v rámci prohlížení rejstříku předmětových kategorií, má pragmatický ráz

a v současné době zahrnuje 97 položek (viz obr. 4/29). Seznam předmětových kategorií je otevřený a ukládající autoři mohou případně navrhnout položky nové. Tvůrci archivu na počátku vymezili dvě základní oblasti, kterým bude věnována mimořádná pozornost: 1. **informační (a počítačové) gramotnosti** a 2. **informetrii (a bibliometrii)** (zejména publikace z výzkumů a doprovodné datové soubory). Aktuální přehled četností obsazených kategorií prozrazuje (viz obr. 4/29 - k 12. prosinci 2008). V archivu se k tomuto datu nacházelo celkem 1377 dokumentů, kterým bylo přiděleno dohromady 179 kategorií (v průměru šlo o necelé 2 kategorie přidělené jednomu dokumentu). Prozatím se v archivu DLIST nalézá největší počet dokumentů věnovaných digitálním knihovnám (186), výchově a vzdělávání v oboru knihovni a informační vědy (141), komunikaci informací ve vědě (134), klasifikacím (112), informační vědě (91) nebo oblasti archivů a WWW (80).

The screenshot shows the DLIST website interface. At the top is a red navigation bar with the DLIST logo and links: Home, Browse, Search, Credits, About, Register, User Area, DL-Harvest, and Help. Below this is a blue header with the text "Digital Library of Information Science and Technology". The main content area has a white background and features the title "Latest News: EPrints Meets Web 2.0" in a large, bold font. Below the title is the author information: "Coleman, Anita Sundaram and Roback, Joseph (2007) 'Latest News': EPrints Meets Web 2.0." A link for the full text is provided: "Full text available as: PDF (22 slides) - Requires Adobe Acrobat Reader or other PDF viewer." An "Abstract" section follows, containing a detailed paragraph about the presentation. At the bottom, there is a metadata section with fields for EPrint Type, Subjects, ID Code, Deposited On, Eprint Statistics, and Tell A Colleague.

DLIST Home | Browse | Search | Credits | About
Register | User Area | DL-Harvest | Help

Digital Library of Information Science and Technology

"Latest News": EPrints Meets Web 2.0

Coleman, Anita Sundaram and Roback, Joseph (2007) "Latest News": EPrints Meets Web 2.0.

Full text available as:
[PDF \(22 slides\)](#) - Requires [Adobe Acrobat Reader](#) or other PDF viewer.

Abstract

This is a presentation at the Open Repositories 2007 Conference, San Antonio, Texas, January 2007. A key Web 2.0 tenet is that users add value and expand the usefulness of the software. EPrints, originally envisioned as software for building a digital repository, is now being extended in many ways by its users. We report on the development of "Latest News" a small feature, we added to our EPrints-2.0 based archive, dLIST. Latest News is wildly popular as a social networking tool with the dLIST communities. dLIST is a disciplinary, cross-institutional archive for the Information Sciences with about 10 editors who connect the fragmented communities in the related areas. It has become obvious that a News module that is more blog-like whereby multiple editors can post News to stay in touch with their respective communities would greatly enhance our efforts to grow active users for the repository. We have now developed Latest News as an EPrints 3.0 plug-in. Scholarly behavior, including self- archiving, varies by discipline but services/features such as News may help all scholars to see themselves as active participants not just in repository growth and use but also its design and software development. The plug-in is available from the Eprints Registry and dLIST and feedback is welcome.

EPrint Type: Presentation

Subjects: [Scholarly Communication](#)
[Software](#)

ID Code: 1782

Deposited On: 20 February 2007

Eprint Statistics: [View statistics for this eprint](#)

Tell A Colleague: [Tell a colleague](#) about it.

Obr. č. 4/30: výstupní prezentační záznam prezentace A. Colemanové ve formátu PPT uložené do repozitáře archivu DLIST [převzato 2008-12-22]

Archiv DLIST nabízí řadu standardních **služeb**. Služby pro ukládající autory byly již zmíněny výše. V rámci **vyhledávání informací** nabízí archiv prioritně prohlížení seznamu **nových přírůstků** za poslední týden. Dále lze vyhledávat informace pomocí rejstříků dokumentů uspořádaných podle roku zveřejnění a podle předmětové kategorie. Předmětovou kategorii si uživatel vybírá klikacím způsobem, smysl má ovšem jenom klikání na položky, které mají uveden alespoň jeden výskyt v databázi indikovaný příslušnou arabskou číslicí. Po získání úplného záznamu (viz obr. 4/30) lze v dalších krocích buď volit úplný text dokumentu, nebo navigovat dále pomocí hyperodkazů na další záznamy (minimálně podle předmětové kategorie). Protože archiv DLIST je provozován ve verzi 2.2, je možné již u některých

záznamů vidět také navigaci ze záznamů citované literatury pomocí odkazu SEEK, stejně jako tomu je i u archivu E-LIS (viz další dílčí část). **Přímé formulování dotazu** lze realizovat buď v jednoduchém režimu (podle slov z názvu, abstraktu a klíčových slov, autorů a editorů a roku zveřejnění) nebo v režimu pokročilém (podle slov z názvu dokumentu, autorů, slov z abstraktu, klíčových slov, předmětových kategorií, typu dokumentu, jména konference, pořadatelské organizace konference, editorů, publikačního statusu, vydavatele a roku zveřejnění).

Registrovaní uživatelé mohou využívat jednak **diskusního listu** archivu (DLIST listserv), jednak mohou získávat **signální služby** (typu Alert) na základě zadaných profilů pomocí elektronické pošty.

Výstupy statistického charakteru patří k velmi sledovaným jak ze strany archivu, tak ze strany archivů konkurenčních, ale i ze strany běžných koncových uživatelů. Statistika využívání archivu DLIST od jeho počátku k lednu 2003 vykazala, kromě jiného, že 6. místo ve využívání obsadili uživatelé z ČR (po USA a Kanadě) [COLEMAN, 2003a, část 8].

Jak bylo uvedeno výše, není archiv DLIST prozatím příliš vyhledávaným archivem. Jeho celkový fond je zatím malý. I to byl zřejmě jeden z důvodů jeho nové plánované inovace (v roce 2004), která mířila k rozšíření archivu o nové moduly, jež by dohromady vytvořily nový systém vyššího řádu nazvaný „**DLISTConnection**“. Koncepce plánovaného systému byla představena v přípravné studii ke grantovému projektu - je uložena v archivu DLIST [COLEMAN, 2003c]. Systém DLISTConnection byl navrhován jako budoucí **služba pro informační a technologickou gramotnost** (ITL, Information and Technology Literacy), která měla významně podpořit gramotnost ve vědě a v oblasti zdravotnictví a která se měla zároveň stát významnou komponentou systému americké Národní vědecké digitální knihovny NSDL (National Science Digital Library, <http://nsdl.org/>). Autoři projektu předpokládali, že k poskytování komplexních služeb systému DLISTConnection bude vytvořen jako jádro speciální digitální fond ITL materiálů, tj. databáze cca 3000 výukových objektů (instruktážní materiály, dokumentace k softwarům, průvodce a manuály k různým bázím a jiné výukové objekty). Předpokládal se jejich přesný metadatový popis. Jak konstatovali autoři projektu, právě tyto typy materiálů nebyly v té době předmětem systematického sběru systému NSDL [COLEMAN, 2003c, I. Purpose]. Pro splnění tohoto cíle bylo nezbytné stávající specifikaci metadat archivu DLIST rozšířit o další prvky. Plánována byla také podrobná specifikace metadat pro potřeby autorskoprávní ochrany získávaných materiálů (řada z nich byla přenesena do systému na základě dohod s komerčními producenty těchto materiálů). Nezbytnou součástí projektu měla být také příprava automatického toku všech informací včetně metadat na základě principů a protokolu OAI-PMH. Testována měla být i interoperabilita archivů provozovaných v různých softwarech (šlo zejména o programy EPrints.org a DSpace). K testování výstupů systému DLISTConnection byly vybrány speciální komunity americké populace - vedle profesionálů působících v oblasti informační a technologické gramotnosti mělo jít zejména o reprezentanty komunity afrických Američanů a původních Američanů, u kterých se předpokládalo velmi citlivé a tudíž účinné vyhodnocování jednotlivých výstupů systému DLISTConnection [COLEMAN, 2003c, VIII. Impact and Expected Results]. Tento plán se k roku 2008 neuskutečnil. Pravděpodobně i místo něho zavedl archiv DLIST novou vyhledávací službu **DL-Harvester** (<http://dlharvest.sir.arizona.edu/>), která je vyhledávacím systémem nad metadaty sklizenými ze 14 archivů v daném oboru a oborech příbuzných (též archivu E-LIS).

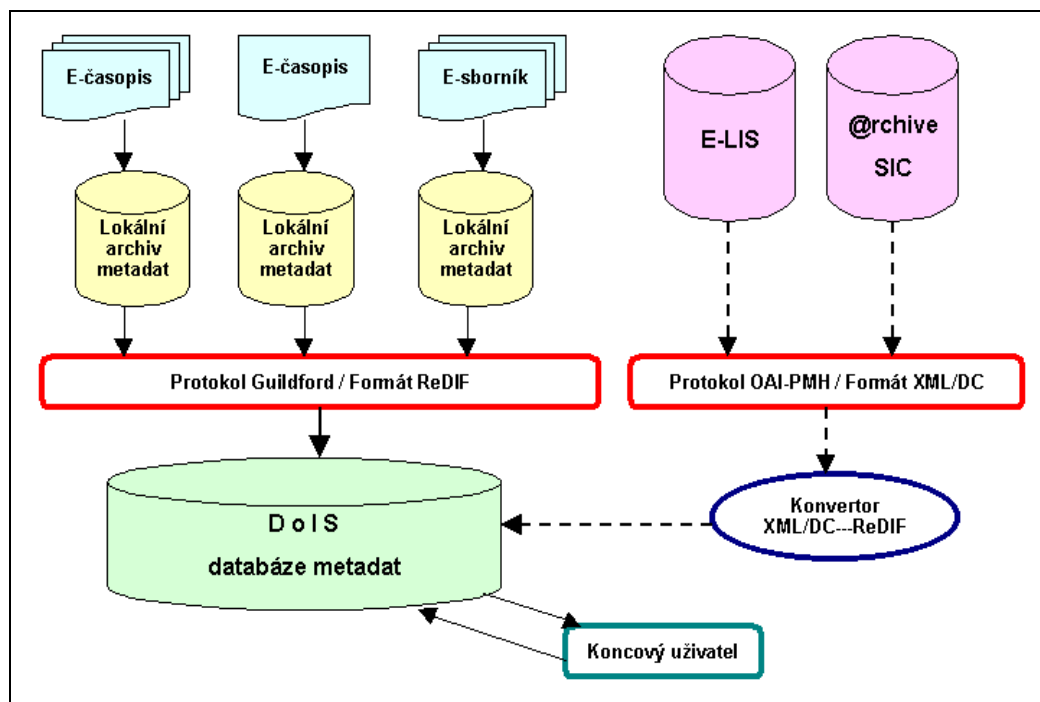
4.3.4 Elektronická knihovna a alternativní bibliografická služba pro informační vědu DoIS

Pod zkratkou „**DoIS**“ se skrývá jméno specifické elektronické knihovny hybridního charakteru pro informace o dokumentech z oboru informační věda a částečně i pro dokumenty samotné, jsou-li k dispozici volně na WWW na mateřských serverech. Je třeba zdůraznit, že jde především o dokumenty pocházející z vědeckovýzkumné činnosti. Plné jméno systému zní „**Dokumenty v informační vědě**“ (Documents in Information Science). O hybridní charakter jde proto, že systém zahrnuje nejen záznamy dokumentů, které jsou volně dostupné v prostoru Internet, ale také záznamy dokumentů, které dostupné volně nejsou (jde zejména o analytické dokumenty z komerčně dostupných časopisů nebo sborníků). Systém bývá označován také termínem „gateway“ nebo „portál“, zejména pak ale termínem „**služba**“. O charakter služby v tomto případě půjde nejvíce a budeme mít na mysli především služby bibliografického charakteru (ve formě komunikace metadat o dokumentech).

Systém byl založen a do provozu uveden již v roce 1999, a to v rámci služby **RCLIS** (Research in Computing, Library and Information Science, <http://rclis.org/>). Služba DoIS je řízena, spravována a udržována mezinárodní skupinou dobrovolníků z různých částí světa. Jde o univerzitní knihovníky, počítačové odborníky a učitele univerzit (také fakult Informační vědy) ze Španělska, Itálie, USA, Indie, Bosny a Hercegoviny aj. Elektronická knihovna je celosvětově volně dostupná a hostuje na serveru Univerzity na Long Islandu (USA, <http://wotan.liu.edu/does/>). Zrcadlové sídlo DoIS je lokalizováno v Itálii (<http://www.dois.it/>).

Hlavní úsilí mezinárodního týmu je vkládáno do tvorby **kvalitních metadat**, především metadat bibliografických jednotek, ukládaných dle stanovené metodiky nejdříve v lokálních archivech registrovaných systémem DoIS. Systém je tedy postaven na základě distribuované architektury (viz obr. 4/31), stejně jako systém RePEc. Metadata, která vytvářejí a ukládají výše zmiňovaní dobrovolníci, jsou z lokálních archivů sklízena pomocí harvestru a na bázi specifického protokolu zvaného „Guildford protocol“ jsou ukládána v centrální databázi systému. Protokol připravil již v roce 1997 hlavní protagonista systému DoIS Thomas Krichel (nyní již ve verzi z roku 2000) [KRICHEL, 2000], který je jinak znám jako jeden z hlavních tvůrců úspěšného systému pro literaturu z ekonomického výzkumu RePEC (viz jeho popis v této kapitole). Bylo to v době, kdy nebyl ještě k dispozici protokol OAI-PMH. V současné době jsou však v rámci systému DoIS konány přípravy i na příjem metadat ze systémů archivů elektronických tisků, které jsou schopny předávat metadata na bázi protokolu OAI-PMH. Na tuto komunikaci je právě připravován i spřízněný archiv E-LIS (viz obr. 4/31). A nic nebrání tomu, aby metadata nebyla sklízena i z oborových archivů dalších (například z archivu @rchiveSIc, jak je znázorněno na obr. 4/31, nebo archivů dalších). Vzhledem k tomu, že samotný systém DoIS na protokol OAI-PMH není připraven, předpokládá se prozatím konverze metadat z formátu XML/Dublin Core do formátu systému DoIS, jímž je formát ReDIF (viz dále v textu). Uvedené plány se prozatím neuskutečnily.

K 12. prosinci 2008 obsahovala databáze systému DoIS **15 183** záznamů článků z elektronických časopisů a **4 500** záznamů příspěvků z odborných akcí (konferencí, zasedání aj.). Nárůst v posledních letech nebyl velký, jedním z důvodů je i fakt, že jeho protagonisté dábají přednost práci v jiných systémech. Převážná část uložených záznamů (14 070) je doplněna hyperodkazem na úplný text dokumentu. Ty jsou uloženy distribuovaně na různých serverech v rámci sítě Internet. V rámci systému se intenzivně provádí i retrospektivní zpracování dokumentů. Pokud jde o konferenční sborníky, bylo plánováno další rozšíření. Zpracovávány by měly být nové tituly sborníků z významných akcí řady mezinárodních organizací.



Obr. č. 4/31: schéma současné (vlevo) a nově plánované (vpravo) architektury systému DoIS [upraveno, zdroj: BARRUECO, 2003a]

Tým dobrovolníků systému DoIS doposud zjišťoval a zpracovával dva základní **typy dokumentů** z oboru informační věda: 1. časopisy a 2. sborníky z odborných akcí. V současné chvíli jsou již také připraveny specifikace pro popis knižních publikací a softwaru. **Popisnou jednotkou** je v převážné míře analytikum, tj. článek z časopisu a příspěvek ze sborníku z akce. U plánovaných knižních titulů se budou popisovat i jejich kapitoly. Záměrem správců systému je celosvětové pokrytí, nicméně již z letmého pohledu na seznam pramenů je patrné, že četně jsou zastoupeny prameny ze Španělska a Itálie (zejména ty, které nabízejí volný přístup k úplným textům), což je dáno také složením týmu dobrovolníků. K jádrovým excerpovaným pramenům patří známé, hojně využívané a volně dostupné časopisy, například časopis *Bulletin des Bibliothèques de France*, *Ariadne*, *D-Lib Magazine* aj. Nicméně, náhled do rejstříku excerpovaných časopisů prozrazuje, že větší množství záznamů reprezentují i články z komerčně dostupných časopisů. Velký počet záznamů pokrýval kolem roku 2004 i produkci prestižního časopisu JASIS/JASIST, který vydává americká společnost ASIS/ASIST. Počet excerpovaných záznamů z časopisu JASIS/JASIST dosahoval počtu téměř 1 000 jednotek. Správci systému se později rozhodli zrušit záznamy z tohoto časopisu, které neodkazují na úplný text, což znamená, že počet záznamů se snížil na 756 (ke konci roku 2008). Některé záznamy již v systému nejsou, například záznam článku z časopisu JASIS autora D. T. Hawkinse, uvedený v publikaci z roku 2003 [BRATKOVÁ, 2003], se již v databázi DoIS nevyskytuje.

Záznamy dokumentů jsou připravovány dle specifikace již výše zmíněného formátu metadat **ReDIF** (Research Documentation Information Format), která vychází z definic známých formulářů IAFA (IAFA templates). Specifikace metadat ReDIF byla původně připravena v rámci výše zmíněného systému pro výzkumné práce a literaturu z oblasti ekonomie RePEc (Research Papers in Economics, <http://repec.org/>), který je jedním z významných účastníků iniciativy propojování elektronických archivů OAI – formát je k dispozici v současnosti již ve verzi z roku 2007 [KRICHEL, 2007]. V záznamech metadat je, vedle jiných údajů (velmi důležitý a cenný je abstrakt v anglickém jazyce), podstatný identifikátor Handle (viz obr

č. 4/32 a 4/33), který je modelován v souladu s požadavky systému DOI (<http://www.doi.org>). Záznam může také obsahovat odkaz vedoucí k získání úplných textů, pokud ovšem nejde o případ komerčně dostupného zdroje (pak adresa na plný text chybí). Na obr. č. 4/32 je pro představu uveden záznam článku z českého časopisu Národní knihovna ve formátu ReDIF, tedy ve struktuře potřebné k uložení v bázi DoIS (nejde tedy o záznam v bázi fakticky uložený).

Template-Type: ReDIF-Article 1.0
Author-Name: Schwarz, Josef
Author-Email: schwarzjv@seznam.cz
Author-Homepage: <http://schwarz.webpark.cz/>
Title: Vybrané teoretické a metodologické problémy terminografie : poznatky z tvorby České terminologické databáze knihovnictví a informační vědy = Selected theoretical and methodological problems of terminography: knowledge acquired in building the Czech terminological database of library and information science
Abstract: An article is based on project called "Czech terminological database of library and information science TDLIS)" that was finished last year (2002). The results of the project are described from theoretical, methodological and historical perspectives. Basic terminographical concepts and theories are shortly outlined, methods and techniques of terminological database compilation are discussed in more details. Key phases are: terminographical team formation, definition of basic project starting points, excerption and identification of concepts and terms and terminological entries compilation. Each step is introduced by description of general methods and concrete examples of TDLIS project solutions are added. History of the Czech and Slovak terminography after 1945 is mentioned and basic information about TDLIS are presented along with overview of TDLIS future developments.
Keywords: Library and information science; Terminology management; Terminology; Terminography; Terminological database; Theory; Methodology; History; Czech Republic; Slovak Republic
Journal: Národní knihovna : knihovnická revue
Pages: 21-41
Volume: 13
Issue: 1
Year: 2003
File-URL: <http://full.nkp.cz/nkkz/Nkkz0204/0204268.html>
File-Format: text/html
File-URL: <http://full.nkp.cz/nkkz/pdf/0301/0301021.pdf>
File-Format: application/pdf
Handle: RCLIS:xxx:nkkrcz:v.13:y.2003:i.1:p.21-41

Obr. č. 4/32: příkladový záznam článku z českého časopisu ve formátu ReDIF (záznam není součástí báze DoIS, jeho identifikátor Handle není reálný) [údaje pocházejí z časopisu Národní knihovna]

Reálný záznam metadat článku publikovaného v časopise JASIST je uveden na obr. 4/33. Byl připraven v jednom z lokálních archivů ve Španělsku, jeho identifikátor je „RCLIS:jul“. Záznamy pro archiv připravuje jeden z dobrovolníků systému DoIS Julio Alonso Arévalo z Univerzity v Salamance [BARRUECO CRUZ, 2003a, část 2.1]. Po bližším prozkoumání údajů záznamů a jejich srovnání s údaji uvedenými na stránkách časopisu JASIS/JASIST je zřejmé, že je zcela využito údajů z amerického serveru, klíčová slova ve španělštině jsou doplněna zpracovatelem.

Vyhledávání informací z databáze metadat systému DoIS má stále jednoduchou povahu. Prvním způsobem vyhledávání je standardní jednoduché **prohlížení rejstříků/seznamů**: 1. zdrojových (excerpovaných) elektronických časopisů a 2. zdrojových (excerpovaných) konferenčních sborníků. V seznamech je v tomto případě dobře vidět, kde jsou a kde nejsou dostupné plné texty.

Druhým způsobem je **přímé formulování rešeršních dotazů** pomocí WWW formuláře rozhraní DoIS. Dotazy je možné formulovat pomocí klíčových slov, která budou prohledávána v celém záznamu metadat. Systém nabízí i hledání v polích Autor (Author) a Název (Title) a Popis (Description). Dotazovací jazyk umožňuje využívání základních booleovských operátorů a možnost rozšíření selekčních údajů pomocí znaku „*“.

Specifickou novou službou systému DoIS je signální služba **RSS** (Rich Site Summary), na základě které jsou uživatelé zpravováni automaticky o obsazích nových čísel volně

dostupných elektronických časopisů v momentě, kdy se objeví na WWW. Prozatím jsou nabízeny články ze čtyř známých časopisů z oboru informační věda: amerického časopisu „D-Lib Magazine“ a tří britských časopisů „Ariadne“, „Information Research“ a „Journal of Digital Information“. Záznamy jsou komunikovány na bázi formátu XML.

The screenshot shows the DoIS website interface. At the top is the DoIS logo and navigation links: DoIS News, Online reference, Conference Proceedings, Journals List, RSS feeds for LIS journals, and About us. Below these is a 'New Search' button. The main content area displays a bibliographic record for an article titled 'An expert system for quality control in bibliographic databases.' by Todeschini, Claudio and Farrell, Michael P. The record includes the journal name, volume, issue, and year (1989, 40(1), 1-11), and a brief abstract. Keywords are listed in Spanish. At the bottom, there is a 'File-URL' and a 'Download Paper' button. The browser status bar at the bottom shows 'Internet' and '100%' zoom.

Obr. č. 4/33: výstupní metadatový záznam článku z komerčního časopisu JASIS uložený v bázi systému DoIS [převzato 2008-12-22]

4.3.5 Závěry

Do světa informačních zdrojů v oboru informační věda vstoupily s intenzivním rozvojem Internetu a informačních technologií dva jejich nové typy: archivy elektronických tisků, které se pokoušejí na jednom místě shromažďovat digitální dokumenty vědeckovýzkumného charakteru, a informační služby hybridního charakteru, jejichž prvořadým úkolem je globální bibliografické mapování oboru informační věda, tedy budování novodobé oborové bibliografie obohacené o odkazy na dostupné úplné texty dokumentů. Oba nové typy silně zasahují a dá se říci i postupně proměňují dosavadní tradiční strukturu informačních zdrojů. Archivy elektronických tisků lze pokládat za novodobé seriály, resp. integrující zdroje, jak jsou dle nejnovější terminologie označovány. V tuto chvíli existují v oboru informační věda již tři archivy a zatím si lze jenom v obrysech představit, kolik jich může přibýt během dalších let, jak budou konkrétně profilovány, jaká bude jejich celková skladba a jak ovlivní komunikaci v oboru.

Analýza čtyř zajímavých systémů ukázala, že v souvislosti s tím, jak se rozvíjely informační technologie, postupně vznikaly jejich jednotlivé typy. Historicky prvním systémem byla

služba DoIS, jejíž představitelé se na bázi zcela dobrovolné pokoušejí vytvořit konkurenci tradičním komerčním bibliografickým a jiným službám z oboru informační věda. Hlavním motivem však nebyla ani tak touha konkurovat, jako využít té situace a skutečnosti, že se v prostředí WWW během desetiletí nakupilo již značné množství literatury, jejíž nezanedbatelná část je volně dostupná. Konečně, komerční bibliografické služby již také delší dobu zpracovávají i volně dostupné dokumenty z Internetu, avšak zpřístupňují je za úplatu, což z jistého pohledu vypadá jako paradox. Bylo by nesmírně užitečné, kdyby i v oboru informační věda vznikla volně dostupná bibliografická služba, která by globálně pokrývala alespoň všechny podstatné volně dostupné i komerční časopisy a periodické sborníky z oboru. Stačilo by k tomu teoreticky málo: opatřit databáze, na základě nichž i komerční vydavatelé zdarma (!) prezentují prostřednictvím záznamů obsahy svých časopisů, nezbytným minimem potřebných technologií a protokolů (OAI-PMH aj.) tak, aby všechny libovolné služby mohly automaticky sklízet metadata a budovat z nich potřebné bibliografické výstupy. Dobrovolníkům ze systému DoIS náleží v tomto prvenství a bylo by velice vhodné tuto službu do budoucna nejen propagovat, ale i napomáhat jejímu zdárnému rozvoji ve prospěch všech uživatelů.

Rok 2000 byl přelomovým rokem v rozvoji tzv. archivů elektronických tisků. Dočkal se i obor informační věda, a tak v prostředí k tomu nejvhodnějším, tedy akademicko-univerzitním, vznikla trojice těchto archivů. Značný vliv na jejich rozvoj měl jednak vznik iniciativy OAI a jejího hlavního produktu - protokolu pro automatický sběr metadat OAI-PMH a budování návazných služeb, jednak intenzivní rozvoj návazných technologií a volně dostupných softwarových produktů pro budování těchto archivů (EPrints.org, DSpace aj.). Dispozice programů a technologií je ovšem jenom předpoklad a počátek budování archivů volně dostupných digitálních dokumentů z výzkumů z oboru informační věda. Nezbytné je archiv jistým způsobem řídit a dále udržovat. Tři analyzované archivy dokladují, že schopný a erudovaný personál, který by se o archivy staral, existuje (jistá dávka nadšení nikde nechyběla). Dá se říci, že provozy těchto archivů jsou novým prvkem celkové informační činnosti univerzit, které by tudíž měly najít minimální prostředky na jejich provoz. Ukazuje se, že se dobře a úspěšně dají takové archivy budovat i v rámci mezinárodní spolupráce (archiv E-LIS).

Ze tří analyzovaných archivů se úspěšně rozvíjejí především evropské projekty. Zajímavý a svébytný francouzský výzkum věd o informacích je pozoruhodně prezentován v archivu @rchiveSIC. Jeho obsah je vysoce kvalitní, bohužel ne všechny příspěvky však bude moci využívat větší počet uživatelů (jazykové bariéry). Nicméně, v archivu se nalézají i dokumenty v anglickém jazyce a podstatné příspěvky francouzských odborníků zaznívají také v angličtině na různých fórech. A tak se pravděpodobně budou dostávat i do druhého evropského archivu E-LIS, který v poslední době značně rozšiřuje objem svého repozitáře i díky iniciativní excerpci příspěvků z řady mezinárodních konferencí konaných na různých místech Evropy i jinde. Pozadu zatím zůstává archiv americký. Pokud se mu však podaří nově koncipovaný záměr rozšíření stávajícího systému DLIST dovést do úspěšného konce, můžeme se dočkat zajímavého digitálního fondu, který bude sloužit v oblasti informační a technologické gramotnosti.

Všechny analyzované archivy úspěšně rozvíjejí dílčí metodické otázky, jako jsou metadatové popisy dokumentů nebo nabízené služby. Daří se rozvíjet další nadstavbové služby, jako je například propojování citovaných a zdrojových dokumentů a budoucí budování citačních rejstříků vystavených nad archivy. Zajímavým problémem archivů z oboru informační věda je věcné pořádání informací. Všechny analyzované archivy přistoupily minimálně k jednoduchému schématu předmětových kategorií, ve kterých se promítá základní problematika i koncepce oboru informační věda ve vybraných zemích či regionech. Ukazuje

se, že předmětové třídění je v archivech nezbytné jak pro generování předmětového rejstříku, který uživatelé rádi navštěvují, tak například pro realizaci signálních služeb (typu Alert). Na příkladu služby DoIS, která nemá zavedeno podobné třídění, je vidět značné ochuzení uživatelů v procesech vyhledávání informací.

4.4 Elektronický archiv CogPrints

Jeden z archivů elektronických tisků, budovaných původně na principu systému/archivu arXiv.org, pro oblast kognitivních věd byl uveden do provozu na Škole elektroniky a počítačové vědy Southamptonské univerzity ve Velké Británii (School of Electronics and Computer Science, University of Southampton, <http://www.ecs.soton.ac.uk/>). Jeho tvůrcem je prof. kognitivních věd **Stevan Harnad**, pod jehož vedením byl archiv, který dostal jméno **CogPrints** (Cognitive Sciences E-Print Archive, <http://cogprints.org/>), po roce 2001 reorganizován za účelem jeho propojení s dalšími archivy a nadstavbovými službami v USA.

Systém je v současné době provozován již na bázi volně dostupného software **EPrints.org** (<http://www.eprints.org/>), nyní v jeho nenovější verzi EPrints 3.1.0 (od června 2007). „CogPrints“ byl nejprve budován s finanční podporou britské Společné komise pro informační systémy JISC (The Joint Information Systems Committee) při Radě pro financování vysokých škol (Higher Education Funding Councils) v rámci známého programu elektronických knihoven „The Electronic Libraries (E-Lib) Programme“. Přesnější tematické vymezení archivu v současnosti zahrnuje následující disciplíny (v závorkách je URL relevantního předmětového rejstříku):

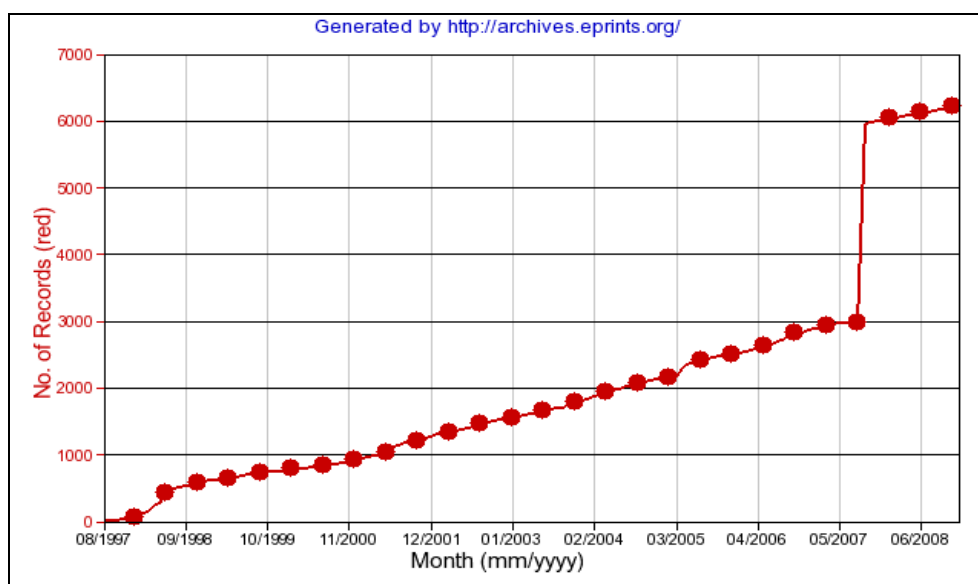
- psychologii (<http://cogprints.org/view/subjects/psyc.html>)
- neurovědu (<http://cogprints.org/view/subjects/neuro.html>)
- lingvistiku (<http://cogprints.org/view/subjects/ling.html>)
- vybrané oblasti počítačové vědy (<http://cogprints.org/view/subjects/comp.html>)
- filozofie (<http://cogprints.org/view/subjects/phil.html>)
- biologie (<http://cogprints.org/view/subjects/bio.html>)

Vstupy archivu jsou zajišťovány podobným způsobem jako v jiných archivech elektronických tisků. Registrovaní autoři ukládají své dokumenty (výzkumné zprávy, konferenční referáty, ale často i články z elektronických časopisů) ve stanovených formátech.

Autoři v rámci ukládání textů připravují a ukládají i metadata (autor, název dokumentu, autorská klíčová slova, abstrakt, předmětová kategorie z nabízeného třídění a případně URL, existuje-li dokument na WWW). Systém využívá vlastní velmi hrubé třídění se 6 základními třídami, které se dále dělí do dílčích skupin.

V repozitáři archivu je, jak uvádí samotný systém v rámci rozhraní pro vyhledávání, ke konci roku 2008 uloženo více než **3 253 dokumentů**, v posledních letech přibývá cca 200 dokumentů ročně. Zřejmě na základě technické chyby uvedl k 15. 12. 2008 registr ROAR v záznamu archivu Cogprints, že zahrnuje již 6 219 dokumentů – příslušná statistika a graf služby Celestial (<http://celestial.eprints.org/>), které jsou zdrojem této informace, tuto chybu zahrnovaly také (viz obr. 4/34).

Do archivu jsou jednotlivě ukládány i starší dokumenty (aktuálně je nejstarším dokument z roku 1950).



Obr. 4/34: graf přírůstků archivu CogPrint [získáno z registru ROAR na základě dat agregátorské služby Celestial 2008-12-15]

V rámci **vyhledávání informací** je možné si prohlížet seznamy záznamů podle věcných kategorií (<http://cogprints.org/view/subjects/>), podle let, nebo je možné formulovat dotaz pro přímé vyhledávání v režimu jednoduchém (Simple, <http://cogprints.org/cgi/search/simple>) nebo pokročilém (Advanced, <http://cogprints.org/cgi/search/advanced>). Selekčními údaji jsou: autor, název dokumentu, klíčová slova abstraktu, rok vytvoření, předmětové třídění, typ dokumentu aj. Nabídka výstupních formátů ke stažení je dána formátem uložení (preferovaným formátem je v současnosti formát PDF, starší dokumenty jsou také uloženy ve formátu Postscript a HTML).

4.5 Předmětové brány systému Intute (dříve RDN a ROADS)

Pod termínem **předmětové brány**, popřípadě **tematické brány (Subject Gateways)** se skrývají moderní digitální systémy, které se v posledních 15 letech cíleně budují a provozují v celosvětovém rámci jako alternativní systémy pro efektivní zpracování, vyhledávání a zejména zpřístupňování vybraných kvalitních volně dostupných internetových informačních zdrojů. V současné době existuje velké množství takových systémů (ve Velké Británii, ve skandinávských zemích, Nizozemsku, Německu, USA, Austrálii a jinde). K základním charakteristikám těchto systémů patří:

- jde o služby založené na **kvalitním** metadatovém **popisu vybraných webových zdrojů** (celých sídel, dílčích celků sídel nebo jednotlivých elektronických dokumentů: elektronických časopisů, diskusních skupin, elektronických konferencí aj.); jednotlivé systémy řeší palčivou problematiku **úrovně popisu i metadat** (údajů o zdrojích), odlišují se tak výrazně od vyhledávacích systémů typu „search engines“, které indexují automaticky každou webovou stránku zvlášť, ale také od univerzálních internetových katalogů, které popisují internetové zdroje úsporným či nedostatečným způsobem
- při přípravě záznamů metadat je nezbytný **vklad ruční lidské práce**, zejména pak odborníků na věcný popis; velká pozornost je proto věnována rozvoji **oborových třídění, řízených slovníků a zejména tezaurů** v dílčích disciplínách nebo tematických oblastech; kvalitní věcný popis je základem úspěšnosti těchto systémů ve srovnání s vyhledávacími systémy současného Internetu

- **vyhledávací systémy** zahrnují jak přímé formulování dotazu prostřednictvím formulářů interaktivního webového vyhledávacího rozhraní, tak pohodlné prohlížení indexů (seznamů) záznamů zdrojů s možností navigace k informačnímu zdroji samotnému
- systémy mají promyšlenou **akviziční politiku** zjišťování a získávání zdrojů, zejména pokud jde o **kritéria výběru** zahrnovaných zdrojů a jejich kvality, jejich hlavním cílem je sloužit vymezené komunitě uživatelů, zejména pak ve vědě, výzkumu a vzdělávání
- budování databáze katalogu internetových zdrojů zahrnuje podporu nutné údržby a aktualizace údajů, zejména pak URL z důvodu někdy značné dynamiky zdrojů informací

Jedním z nejznámějších systémů, které se v rámci tohoto typu digitálních systémů již dlouhou rozvíjejí, je systém s aktuálním označením **Intute**, který se postupně zrodil na bázi předchozího systému a sítě RDN (viz dále v textu) a ještě před ním na bázi předmětových bran budovaných v původním systému a softwaru ROADS. V další části této podkapitoly jsou nejdříve uvedeny stručné informace o předchůdcích systému Intute, v části další pak podrobnější informace o samotném systému Intute.

4.5.1 Počátky - předmětové brány v systému ROADS

Počátky výše zmíněné Sítě pro vyhledávání zdrojů RDN (Resource Discovery Network) a později systému Intute jsou spojeny s významným britským programem „**E-Lib**“ (**Electronic Libraries Programme**, <http://www.ukoln.ac.uk/services/elib/>), který byl od roku 1996 grantován Společnou komisí pro informační systémy JISC (Joint Information Systems Committee, <http://www.jisc.ac.uk/>), jež působí při Nadační radě pro vysokoškolské vzdělávání ve Spojeném Království Velké Británie a Severního Irska. Šlo o široce založený program, který řešil řadu úloh (digitalizaci zdrojů, elektronické publikování, elektronické online časopisy, elektronické dodávání dokumentů aj.). Na první fázi řešení, která trvala tři roky, navázala fáze druhá, která trvala až do roku 2001. Program „E-Lib“ byl finančně podpořen také Evropskou Unií, a to v rámci jiného ale souvisejícího programu **DESIRE** (<http://www.desire.org/>), který byl realizován v letech 1996-2000 (řešeny byly zejména otázky zpracování a zpřístupňování internetových zdrojů pomocí předmětových bran a metadatový popis). Jiným programem, financovaným v dalších letech Evropskou Unií, byl program **RENARDUS** (<http://www.ukoln.ac.uk/metadata/renardus/>). Byl zaměřen na budování kvalitního zpřístupňování internetových zdrojů v oblasti vědy, výzkumu a vzdělávání.

Jedním z významných výsledků programu „E-Lib“ byl vznik řady předmětových bran pro zpracování a zpřístupňování internetových zdrojů, které byly vybudovány v rámci dílčích projektů v programovém vybavení **ROADS (Resource Organization And Discovery in Subject-based services)**. Předmětové brány byly formálně seskupeny v rámci sítě **ANR (Access to Network Resources)** [HIOM, 2006]. (Samotné programové vybavení systému ROADS vzniklo také v rámci samostatné části výzkumného programu „E-Lib“ (<http://www.ukoln.ac.uk/metadata/roads/>); systém zahrnoval také definovaný standard metadat ROADS včetně typologie informačních zdrojů). K významným předmětovým branám v systému ROADS patřily (adresy URL některých systémů jsou již automaticky přesměrovávány na části systému Intute):

- **SOSIG** : Social Science Information Gateway (dřívější URL: <http://sosig.ac.uk/>); brána SOSIG byla jako pilotní program dokonce připravena jeden rok před zahájením výzkumného programu „E-Lib“ v roce 1995 s finanční podporou britské Rady pro ekonomický a sociální výzkum ESRC (Economic and Social Research Council); SOSIG se proměnil v dnešní systém Intute pro společenské vědy

- **ADAM** : Art, Design, Architecture & Media Information Gateway (dřívější URL: <http://www.adam.ac.uk>); ADAM se proměnil v dnešní systém Intute pro umění a humanitní vědy
- **RUDI** : Resource for Urban Design Information (<http://www.rudi.net/>)
- **OMNI** : Organising Medical Networked Information (dřívější URL: <http://www.omni.ac.uk/>); OMNI se proměnil nejdříve v systém BIOME a později v dnešní systém Intute pro vědy o zdraví a o životě
- **BIZ/ED** : Biz/ed: Business Education on the Internet (dřívější URL: <http://www.bized.co.uk/>)
- **EEVL** : Edinburg Engineering Virtual Library (dřívější URL: <http://www.eevl.ac.uk/>); EEVL se proměnil v dnešní systém Intute pro přírodní vědy, inženýrství a technologie

Systém **ROADS** byl konkrétně rozvíjen ve spolupráci tří institucí: **Fakulty počítačové vědy na Loughboroughské technické univerzitě** (Department of Computer Science, Loughborough University of Technology), **Institutu pro výukové a výzkumné technologie na Bristolské univerzitě** (Institute for Learning and Research Technology, Bristol University) a **Úřadu pro knihovnické a informační sítě ve Spojeném Království UKOLN při Bathské univerzitě** (UK Office for Library and Information Networking, University of Bath, <http://www.ukoln.ac.uk/>). Programový systém byl připraven v jazyce PERL a pracoval pod OS UNIX (LINUX aj.). **Metadata** byla vyvážena pomocí formulářů, které byly modifikací formulářů navržených skupinou **IAFA** (Internet Anonymous FTP Archive Working Group), jež patří pod organizaci **IETF** (Internet Engineering Task Force). **Věcný popis** byl v každém jednotlivém systému různý, řada systémů převzala a rozvíjela známé tezaury z klasických tradičních dokumentografických systémů (MeSH, AGRIS, aj.). **Vyhledávání informací** bylo v zásadě podobné ve všech systémech (prohlížení a navigování nebo formulace dotazu v rámci rozhraní pro vyhledávání). Některé systémy měly programově dopracovány vazby na tezaurus, který bylo možné využívat při vyhledávání (SOSIG).

4.5.2 Síť předmětových bran RDN

Jednotlivé výše uvedené systémy předmětových bran (v průběhu času přibýly také další brány) byly velkým úspěchem celého programu „E-Lib“. V zájmu jejich **dalšího rozvoje** navrhla již v roce 1998 národní komise JICS jejich sdružení do národní **Sítě pro vyhledávání zdrojů RDN (Resource Discovery Network)**. Komise JISC ve spojení s britskou Radou pro ekonomický a sociální výzkum ESRC (Economic and Social Research Council) a také Radou pro výzkum v oblasti umění a humanitních věd AHRC (Arts and Humanities Research Council) garantovala financování rozvoje. V lednu 1999 bylo zřízeno centrum sítě **RDNC (Resource Discovery Network Centre)** na kterém se partnersky podílely King's College v Londýně a úřad UKOLN s podporou Hullské univerzity (University of Hull).

Některé z výše uvedených předmětových bran se v rámci nové sítě začaly spojovat s dalšími institucemi, které působily v relevantní předmětové oblasti. Například původní systém **OMNI** se spojil s Muzeem pro historii přírody (Natural History Museum), Readingskou univerzitou (University of Reading), specializovanou knihovnickou komisí BRIL (BBSRC Research Institutes Librarians Committee) a Královskou kolejí veterinárních chirurgů (Royal College of Veterinary Surgeons) a vytvořily společně bránu (informační službu) zvanou **BIOME**. Významný systém **SOSIG** (SOSIG byl v podnázvu rozšířen o oblast ekonomických věd a právo) začal spolupracovat s Britskou knihovnou pro politické a ekonomické vědy (British Library for Political and Economic Science), Výpočetním centrem pro sociální vědy a antropologii Canterburské univerzity v Kentu (Centre for Social and Anthropological

Computing at the University of Canterbury at Kent), Surreyskou univerzitou (University of Surrey), Londýnskou univerzitou a dalšími institucemi. Vznikla také nová brána **MathGate Project** pro inženýrské obory, matematiku a počítačovou vědu, kterou spravovala Birminghamská univerzita (University of Birmingham).

Ke konci roku 2005 bylo do sítě RDN zahrnuto celkem asi 70 partnerských univerzit a dalších výzkumných institucí [HIOM, 2006]. K předmětovým branám sítě RDN náleželo celkem 8 systémů:

1. Altis : Hospitality, Leisure and Tourism (původní URL: <http://www.altis.ac.uk/>); dnes je součástí systému Intute pro společenské vědy
2. Artifact : Arts and Creative Industrie (původní URL: <http://www.artifact.ac.uk/>); dnes je součástí systému Intute pro umění a humanitní vědy
3. BIOME : Health and Life Sciences (původně OMNI) (původní URL: <http://biome.ac.uk/>); dnes je součástí systému Intute pro vědy o zdraví a o životě
4. EEVL : Engineering, Math, and Computing (původní URL: <http://www.eevl.ac.uk/>); dnes je součástí systému Intute pro přírodní vědy, inženýrství a technologie
5. GEsources : Geography and Environment (původní URL: <http://www.gesources.ac.uk/>); dnes je součástí systému Intute pro přírodní vědy, inženýrství a technologie
6. Humbul : Humanities (původní URL: <http://www.humbul.ac.uk/>); dnes je součástí systému Intute pro umění a humanitní vědy
7. PSIGate : Physical Science (původní URL: <http://www.psigate.ac.uk/>); dnes je součástí systému Intute pro přírodní vědy, inženýrství a technologie
8. SOSIG : Social Sciences, Business and Law (původní URL: <http://sosig.ac.uk/>); dnes je součástí systému Intute pro společenské vědy

Správy celé sítě RDN se ujala Manchesterská univerzita (University of Manchester, <http://www.manchester.ac.uk/>), technickou podporu zajišťoval UKOLN a výše uvedených 8 služeb předmětových bran (všechny jsou v současné době již součástí **systému Intute**).

Jedním z cílů sítě RDN bylo spojit úsilí a zkušenosti expertů, které v průběhu celého vývoje jednotlivé předmětové brány postupně získaly. Centrum sítě formulovalo a kodifikovalo novou politiku získávání, zpracování a zpřístupňování internetových zdrojů pro potřeby výzkumu a vzdělávání ve Spojeném Království. Byly připraveny nové směrnice a příručky. Byly zpřesněny toky všech informací v celém systému, zejména výměna metadatových záznamů mezi organizacemi. Výsledkem dílčího výzkumu byla příprava technického projektu pro zavádění a údržbu **předmětových portálů RDN (RDN Subject Portals Project (SPP))**, který měl v dalším období umožňovat optimální vyhledávání napříč všemi systémy předmětových bran. Portálová řešení bran byla testována a zaváděna od roku 2004.

Síť RDN od roku 1999 ve své hlavní bázi katalogu nashromáždila cca **120 000 metadatových záznamů internetových zdrojů**. Kromě hlavní služby vyhledávání informací v rámci webových rozhraní byly zavedeny i další návazné služby (například služby pro potřeby školení uživatelů v jednotlivých oborech, personalizované přístupy do systémů bran aj.).

V roce 2005 byla na základě rozsáhlé analýzy systémů předmětových bran v měnícím se prostředí sítě Internet a služby WWW (nastoupil, kromě jiného, i směr Web2.0) a také interních konzultací odborníků sítě RDN zahájena **nová fáze rozvoje předmětových bran**. Byla zahájena **proměna** této sítě do **nového systému Intute**. Do hry vstoupila restrukturalizace předmětových služeb, integrace softwarového zabezpečení a hardwarových

platformem a také nová vizuální identifikace těchto služeb. Po intenzivních přípravách (18 měsíců) byl nový systém Intute spuštěn 7. 7. 2006 (viz dále v textu) [WILLIAMS, 2006].

4.5.3 Systém předmětových bran Intute

Systém Intute (<http://www.intute.ac.uk/>) je volně dostupnou **online informační službou**, jejímž cílem je zpracování a zpřístupňování vysoce kvalitních informací o online dostupných internetových zdrojích, které se vztahují k **výuce, vzdělávání a výzkumu** v rámci oborů pěstovaných na vysokých školách a v rámci kursů celoživotního vzdělávání ve Spojeném Království. Informace jsou zpřístupňovány prostřednictvím předmětově (oborově) profilovaného katalogu. Systém Intute nezpracovává a nezpřístupňuje žádné klasické tištěné informační zdroje. Informační fondy systému Intute jsou přísně výběrové, zahrnovány jsou pouze zdroje, které odpovídají stanoveným **výběrovým kritériím**. Stavebním kamenem je **kvalita zdrojů** posuzovaná co do hodnoty profesionály ze zahrnutých vědních disciplín [BURGESS, 2006, kap. 2.2]. **Metadatové záznamy** jsou dostupné zdarma všem návštěvníkům internetu, stejně jako informační zdroje samotné (až na určité výjimky materiálů a zdrojů dostupných na bázi licence).

Celý systém Intute pracuje na bázi konsorcia (Intute Consortium), složeného z velkého množství partnerských institucí (jsou vyjmenovány níže v seznamu subsystémů systému Intute). Systém má nyní **jedinou identitu** a má **jedno společné webové uživatelské rozhraní**. Organizační struktura celého systému zahrnuje:

- **Výkonný výbor Intute** (Intute Executive Committee), který je zodpovědný za strategický rozvoj celého systému a jeho celkové řízení. Práce Výkonného výboru je zajišťována informační službou MIMAS (Manchester Information & Associated Services, <http://www.mimas.ac.uk/>), která patří pod Manchesterskou univerzitu (The University of Manchester)
- Čtyři **Předmětové skupiny** (Subject Groups), které zabezpečují konkrétní služby ve vymezené předmětové oblasti včetně **dílčích částí databáze katalogu systému Intute**

Jádrem celého systému **Intute** (<http://www.intute.ac.uk/>) je tedy databáze **katalogu všech internetových zdrojů (Internet Resource Catalogues, IRCs)**, kterou provozuje a udržuje informační služba MIMAS (<http://www.mimas.ac.uk/>). Zahrnuje metadatové záznamy (k polovině prosince 2008 je jich cca 125 000), které jsou vytvářené decentralizovaně v rámci čtyř předmětových (tematických) subsystémů:

- **Intute : Umění a humanitní vědy** (Intute : Arts and Humanities, <http://www.intute.ac.uk/artsandhumanities/>), zajišťuje Oxfordská univerzita (University of Oxford) ve spolupráci s Manchesterskou metropolitní univerzitou (Manchester Metropolitan University)
- **Intute : Vědy o zdraví a o životě** (Intute : Health and Life Sciences, <http://www.intute.ac.uk/healthandlifesciences/>), zajišťuje Nottinghamská univerzita (University of Nottingham)
- **Intute : Přírodní vědy, inženýrství a technologie** (Intute: Science, Engineering and Technology, <http://www.intute.ac.uk/sciences/>), zajišťuje Manchesterská univerzita (University of Manchester) ve spolupráci s Univerzitou Heriota Watta (Heriot Watt University)
- **Intute : Sociální vědy** (Intute : Social Sciences, <http://www.intute.ac.uk/socialsciences/>), zajišťuje Bristolská univerzita (University of Bristol) ve spolupráci s Birminghamskou univerzitou (University of Birmingham)

Bristolská univerzita navíc zajišťuje školící služby pro celý systém Intute (Intute training services). Systém Intute je dnes financován Společnou komisí pro informační systémy JISC (Joint Information Systems Committee). Subsystem umění a humanitních věd je částečně dotován také Radou pro výzkum v oblasti umění a humanitních věd AHRC (Arts and Humanities Research Council).

Základní příručka systému Intute definuje následující **typy a kategorie zpracovávaných internetových zdrojů** [BURGESS, 2006, kap. 3.6]:

- Výzkumné projekty a zprávy (pocházející zejména z národních britských nadací)
- Výukové a vzdělávací projekty a jejich zprávy
- Služby a zdroje zajišťované Společnou komisí pro informační systémy JISC (Joint Information Systems Committee), které jsou předmětově relevantní systému Intute
- Specifické služby zajišťované v rámci systému Intute:
 - Virtuální školící či výukové materiály
 - Relevantní případové studie pro projekty celoživotního vzdělávání
 - Další služby s přidanou hodnotou nabízené systémem Intute a jeho dílčími subsystemy
- Bibliografie a bibliografické databáze
- Elektronické časopisy:
 - Úplné texty časopisů, zejména volně dostupných a recenzovaných (lektorovaných)
 - Seznamy abstraktů a obsahů časopisů
- Seznamy e-mailových adres a diskusních skupin, zejména pokud mají webové archivy
- Domovské stránky (webová sídla) akademických institucí Spojeného Království, které působí v oborech relevantních zaměření systému Intute
- Výukové a vzdělávací materiály včetně online příruček
- Adresáře informačních zdrojů, zejména zdrojů s originálním obsahem nebo formou
- Audiovizuální zdroje, videozáznamy a jiné multimediální zdroje
- Webová sídla se seznamy aktuálních událostí (akcí)
- Webová sídla relevantních organizací a asociací (orgánů státní správy, profesionálních institucí a vybraných spolupracujících komerčních společností)
- Referenční materiály včetně encyklopedií a slovníků
- Náborové a zaměstnanecké služby
- Služby v oblasti médií včetně novin, online zpravodajských serverů a televizních stanic
- Online databáze včetně služeb na předplatné, které jsou považovány za užitečné a které jsou snadno dostupné v rámci sektoru vysokoškolského a celoživotního vzdělávání ve Spojeném Království
- Knihovny, muzea a archivy, které v rámci svých webů nabízejí významné informační zdroje vztahující se k jejich sbírkám, výstavám aj.
- Primární zdroje v elektronické formě (textové, obrazové, zvukové aj.)
- Sekundární zdroje v elektronické formě bez rozlišení, jde-li o zdroje publikované přímo na webu nebo zdroje digitalizované z původních zdrojů tištěných

Úplný seznam všech typů zpracovávaných zdrojů a jejich definice je uveden v rozsáhlé příloze č. 2 základní příručky systému Intute [BURGESS, 2006, příl. č. 2].

Ze zpracování jsou v systém Intute **vyloučeny** zdroje, které podporují nelegální činnost, zdroje, které porušují britské zákony, zdroje, které nesplňují kritéria výběru, jednoduché

seznamy hyperlinků na informační zdroje s krátkými anotacemi nebo bez nich, zdroje určené pro marketing, reklamu a komerční produkty, domovské stránky fakult univerzit, které nezahrnují významné online zdroje, webové stránky akademických pracovníků, které neobsahují významné online zdroje a personální stránky jednotlivců, které sestávají pouze z informací o této osobě.

V systému Intute nejsou zavedena žádná geografická omezení, pokud jde o původ informačních zdrojů, zvláštní důraz je však kladen na **zdroje pocházející přímo ze Spojeného Království**. Zdroje ze zahraničí jsou zahrnovány výběrově, když jsou vyhodnoceny jako vhodné. Prioritně systém zpracovává informační zdroje v angličtině nebo zdroje do angličtiny přeložené z jiných jazyků. Zahrnovány však jsou i zdroje v dalších jazycích, jejich popis je však veden na bázi angličtiny.

Všichni uživatelé systému Intute mají úplný přístup k veškeré dokumentaci a příručkám systému a mohou kdykoliv navrhnout nové informační zdroje ke zpracování prostřednictvím webového formuláře (<http://www.intute.ac.uk/suggest.html>). Příspěvatelé a katalogizátoři systému Intute procházejí předepsaným školením. Informační zdroje jsou vyhledávány v prostoru Internetu a WWW různými metodami (kromě metody návrhu na zpracování zdroje od běžných uživatelů se využívá zejména známých vyhledávacích systémů, internetových bran, blogů, diskusních skupin apod.). Informační zdroje vybírají odborníci (experti) z oborů zahrnutých v systému Intute. K **základním kritériím** výběru zdrojů patří užitečnost zdroje pro výzkum, výuku a vzdělávání zejména s ohledem na potřeby uživatelů ve Spojeném Království. K **dalším kritériím** patří autorizovanost a důvěryhodnost zdroje, přesnost uváděných faktů, originální obsah zdroje, jeho objektivnost, citovanost, aj. a v neposlední řadě také formální vzhled zdroje a jeho vybavení.

Katalogizace zjištěných a vybraných informačních zdrojů se řídí předepsanými pokyny, které jsou dostupné v katalogizační příručce dostupné ze webových stránek systému (http://www.intute.ac.uk/intute_cataloguing_guidelines_v4_1.doc). Popis je založen především na standardu **Dublin Core** a standardu pro metadatový popis **RLLOMAP**. (<http://www.intute.ac.uk/publications/rdn-ltsn-ap/>). Využívána jsou ale také katalogizační pravidla **AACR2** a přihlíží se již také k návrhu nového mezinárodního standardu pro popis zdrojů **RDA** (Resource Description and Access). Pro předmětový popis se využívá specifických příruček pro jednotlivé zahrnuté obory. Systém Intute spravuje a rozvíjí **tezaurus deskriptorů** (<http://www.intute.ac.uk/socialsciences/thesaurus.html>). Vyhledávání v databázi tezauru je propojeno s vyhledáváním v metadatové databázi. Každý metadatový záznam získává jedinečný identifikátor.

V rámci **správy a údržby** celého informačního fondu systému Intute dochází k případným aktualizacím vytvořených záznamů. Záznamy zastaralých zdrojů mohou být (pouze logicky) vymazány a nahrazeny záznamy zdrojů nových, metadata mohou být v případě potřeby aktualizována (nutností je zejména kontrola a oprava neplatných URL adres, specifický program „Website Watcher“ automaticky hlídá změny realizované v rámci registrovaných webových sídel). Zbytečné duplicity záznamů stejných zdrojů jsou likvidovány, jisté množství zdrojů je ale v celém systému zpracováno záměrně duplicitně napříč 4 zahrnutými subsystemy. Důležitým principem správy systému je udržování optimálního počtu záznamů (v současnosti jde o fond cca 125 000 metadatových záznamů, jak bylo uvedeno výše). Velký nárůst by mohl znamenat snížení kvality a úrovně zpracování dat.

Základní služba systému Intute, tj. volné **vyhledávání informací**, je zajišťována přes jednotné webové rozhraní, které je v základní formě, tj. **jednoduchém vyhledávání** (Search), nabízeno přímo z hlavní stránky celého systému (<http://www.intute.ac.uk>). Vyhledávání může

být realizováno buď v rámci celé databáze metadatových záznamů (katalogu) nebo v některé jeho předmětové části.

Z hlavní stránky systému Intute je uživatelům nabízeno také pohodlné vyhledávání záznamů pomocí prohlížení **rejstříku hlavních předmětových kategorií** (Browse) nebo podrobného rejstříku předmětových hesel (Browse Headings A-Z), který zároveň u hesel zobrazuje počty relevantních záznamů informačních zdrojů.




Obr. 4/35: rozhraní pro vyhledávání informací v celém systému Intute s formulací dotazu a zobrazením prvního zkráceného záznamu zdroje z části pro humanitní obory [cit. 2008-12-18]


Z hlavní stránky systému Intute se lze také dostat k seznamu zcela **nových záznamů** (New resources). Specifickou nabídkou je vyhledávání v části databáze katalogu systému Intute se zaměřením na informační zdroje vhodné pro **celoživotní vzdělávání** (Resource for FE, tj. Further Education, <http://www.intute.ac.uk/fe/index.html>).


K dispozici je pro uživatele také **pokročilé vyhledávání** (Advanced search, <http://www.intute.ac.uk/search.html>) – v jeho rámci lze volit 3 specifické selekční údaje (název zdroje, popis zdroje a klíčová slova). Bez volby selekčního údaje systém automaticky prohledává celý metadatový záznam (All fields). Formulace dotazu může být navíc limitována údajem o typu zdroje (Resource type(s)). Uživatel může v případě pokročilého vyhledávání také předem zvolit způsob uspořádání vyhledaných záznamů (podle relevance nebo podle abecedy názvů zdrojů). Vybrat lze také formát zobrazení záznamů (všechny údaje nebo jen názvy zdrojů).

Z obrazovky pokročilého vyhledávání lze přejít pomocí odkazu „**Intute Harvester**“ k vyhledávání v širší databázi záznamů internetových zdrojů, která zahrnuje navíc záznamy zdrojů nepracovaných ručním způsobem (jsou sklizeny z webu automatickou cestou). Tuto možnost lze využít, pokud uživatel není spokojen s výsledky z vyhledávání v základním katalogu. Nicméně, vyhledávací systém systému Intute nabízí vždy po každé rešerši (i v rámci jednoduchého rozhraní) ve fázi zobrazování seznamů nalezených zkrácených záznamů hyperlink (Harvester results), který vede k vedlejší automaticky připravené rešerši záznamů ze širší databáze na daný dotaz (viz obr. 4/35).

Zobrazování metadatových záznamů zkrácených i úplných je v systému Intute na vysoké úrovni též z hlediska grafiky. Uživatel se může orientovat podle následujících ikon (jsou vždy jako nápověda přítomny na začátku seznamu nalezených zkrácených záznamů), které jsou zobrazovány u záznamů:

 Záznamy ze subsystému: Intute : Science, Engineering and Technology

 Záznamy ze subsystému: Intute : Arts and Humanities

 Záznamy ze subsystému: Intute : Social Sciences

 Záznamy ze subsystému: Intute : Health and Life Sciences

 Uložení záznamu

Úplné metadatové záznamy jsou velmi rozsáhlé. Zahrnuty jsou jak podrobné formální údaje, tak zejména podstatné údaje věcné povahy (klíčová slova-deskriptory u tezaurů, existují-li v daném oboru, neřízená klíčová slova, popis (abstrakt) a oborová třídění). Podstatným údajem je lokalizace zdroje v síti Internet (URL). Jednotlivé údaje jsou označeny návěštními v angličtině a řada údajů je v systému hyperlinkována (viz obr. 4/36).

Systém Intute nabízí kromě vyhledávacích služeb některé další specifické služby, například:

- Službu MojeIntute (MyIntute, <http://www.intute.ac.uk/myintute/index.php>) k personalizovanému přístupu do systému Intute
- Službu Informs (http://www.informs.intute.ac.uk/informs_perl/login2.pl), která umožňuje interaktivní tvorbu výukových pomůcek

Další služby a související systémy jsou připravovány.

Jednou z nich je také perspektivní vyhledávací služba „**Intute Repository Search**“ (<http://www.intute.ac.uk/irs/>), která je právě v období testování. Jde o vyhledávací systém nad metadatovou databází, která obsahuje záznamy dokumentů uložených v 95 akademických repozitářích Velké Británie. Fond zahrnuje již **413 474 dokumentů** (pracovních studií, výzkumných zpráv, článků z časopisů, materiálů z konferencí, a jiných vědeckých dokumentů). Projekt je realizován v rámci aktivit komise JISC (od roku 2006).

Subject links About us A-Z of services Internet catalogue Search Browse New resources Suggest a site Internet training Virtual Training Suite Support materials Intute events Additional Services Blog E-journals search MyIntute Newsround Science data Spotlight Timelines World guide	Title	CiteSeer scientific literature digital library and search engine <small>Editor's Choice</small>
	Also known as	CiteSeerX
	Description	Full text indexed computer and information science research articles and papers can be searched from this database. Citations made by indexed documents are also listed. Other features available are lists of the most cited articles and authors. Documents are available in a variety of formats. The CiteSeer database was developed by the NEC Research Institute and is now hosted at the Pennsylvania State University's College of Information Sciences and Technology.
	Keywords - uncontrolled	intelligent agents ; computer architecture ; artificial intelligence ; expert systems ; optimization ; compression ; hardware ; human computer interaction ; multimedia ; information retrieval ; digital libraries ; neural networks ; operating systems ; programming ; security ; software engineering ; data structures ; parallelism ; computational computing
	Type	Bibliographic databases; Datasets; Papers/reports/articles (collections)
	URL	http://citeseerx.ist.psu.edu/ [English;]
	URL	http://citeseerx.ksu.edu.sa/ [English;]
	Classification	Computing > Computing General
	Publisher	College of Information Sciences and Technology, Pennsylvania State University
	Country of origin	United States

Obr. 4/36: úplný metadatový záznam inovované digitální knihovny CiteSeerX z fondu subsystému Intute pro přírodní vědy, inženýrství a technologie [cit. 2008-12-18]

4.6 Volně dostupné digitální knihovny s citačními službami

4.6.1 Úvod

Narůstající fondy vědeckých informačních materiálů, které jsou dostupné zdarma v prostoru WWW, jsou předmětem zájmu nejen koncových uživatelů, kteří je stále více vyhledávají a využívají ve své práci, ale také předmětem značného zájmu informačních profesionálů a řady informačních služeb, které tyto materiály zpracovávají a dále zpřístupňují ve formě zajímavých výstupů. Ke službám velkého významu v oblasti vědy a výzkumu patří služby citační, které, kromě jiného umožňují sledovat vliv vědeckých prací na další vývoj vědeckého výzkumu. Text této podkapitoly je věnován stručnému rozboru a hodnocení vybraných systémů (volně dostupných), které zpracovávají volně dostupné materiály, uložené na internetu nebo v digitálních knihovnách či elektronických archivech, a také a citace v nich obsažené. Text vychází z informací získány prostřednictvím online rozhraní jednotlivých citačních systémů a také z dostupné literatury k tomuto tématu [HARDY, 2005]. Text nezahrnuje komerční systémy, které také již zpracovávají volně dostupné materiály (WOS, Web Citation Index, Scopus a CrossRef) – jejich popis a zhodnocení však autorka zveřejnila v obsáhlejší konferenčním příspěvku [BRATKOVÁ, 2006b]. Představeny jsou systémy volně dostupných citačních služeb: Gogole Scholar, CiteSeer/CiteSeerX a Citebase Search.

4.6.2 Citační indexace a citační služby

Citační indexaci rozumíme proces vytváření rejstříku citací. **Citační rejstřík** je databázi, která spojuje citující dokumenty s dokumenty citovanými prostřednictvím jejich záznamů. Zatímco seznam záznamů použité (citované) literatury v nějakém dokumentu míří vždy jen k **dříve publikovaným dokumentům**, specifickým citačního rejstříku je, že může poskytnout také seznamy **později publikovaných dokumentů**, které daný dokument citovaly (vazba typu „cited by“). Citační rejstříky jsou významnými nástroji citační analýzy, v rámci které jsou sledována citovaná díla za účelem určení vlivu těchto děl nebo jejich autorů. Citace mohou být využity k měření **vlivu** nějakého dokumentu ve specifickém oboru. Čím více je čten a citován nějaký článek, tím více může ovlivňovat badatele v tomto oboru. Na úrovni citačních služeb je měřen tzv. „**impakt faktor**“ (faktor vlivu) vědeckých časopisů.

Se zdokonalováním informačních technologií se citační rejstříky staly sofistikovanějšími. Oproti dřívějšímu období vstoupily do hry **úplné texty dokumentů** v dobře **organizovaných digitálních knihovnách nebo archivech** s předplaceným nebo otevřeným přístupem. Takové prostředí umožňuje poskytování velkého počtu nových zajímavých a užitečných výstupů a služeb. Jsou-li připravena kvalitní hypertextová propojení, může uživatel získávat úplné texty jak citujících tak citovaných dokumentů. Zdroje s otevřeným přístupem (archivey a volné online časopisy), komentované v předchozí kapitole, jsou zahrnovány dnes i do komerčních citačních služeb, které se tak stávají ještě více komplexnějšími. S nástupem nových technologií skončila zároveň dlouhá doba monopolního postavení zatím největšího a nejkvalitnějšího světového citačního rejstříku na světě provozovaného dnes v systému **Web of Science (WOS)**, který je součástí komplexního portálu **Web of Knowledge**. Objevily se a objevují i nadále další nové moderně koncipované citační služby, které se na nekomerční bázi věnují informačním materiálům s otevřeným přístupem.

4.6.3 Vybrané volně dostupné systémy s citačními službami

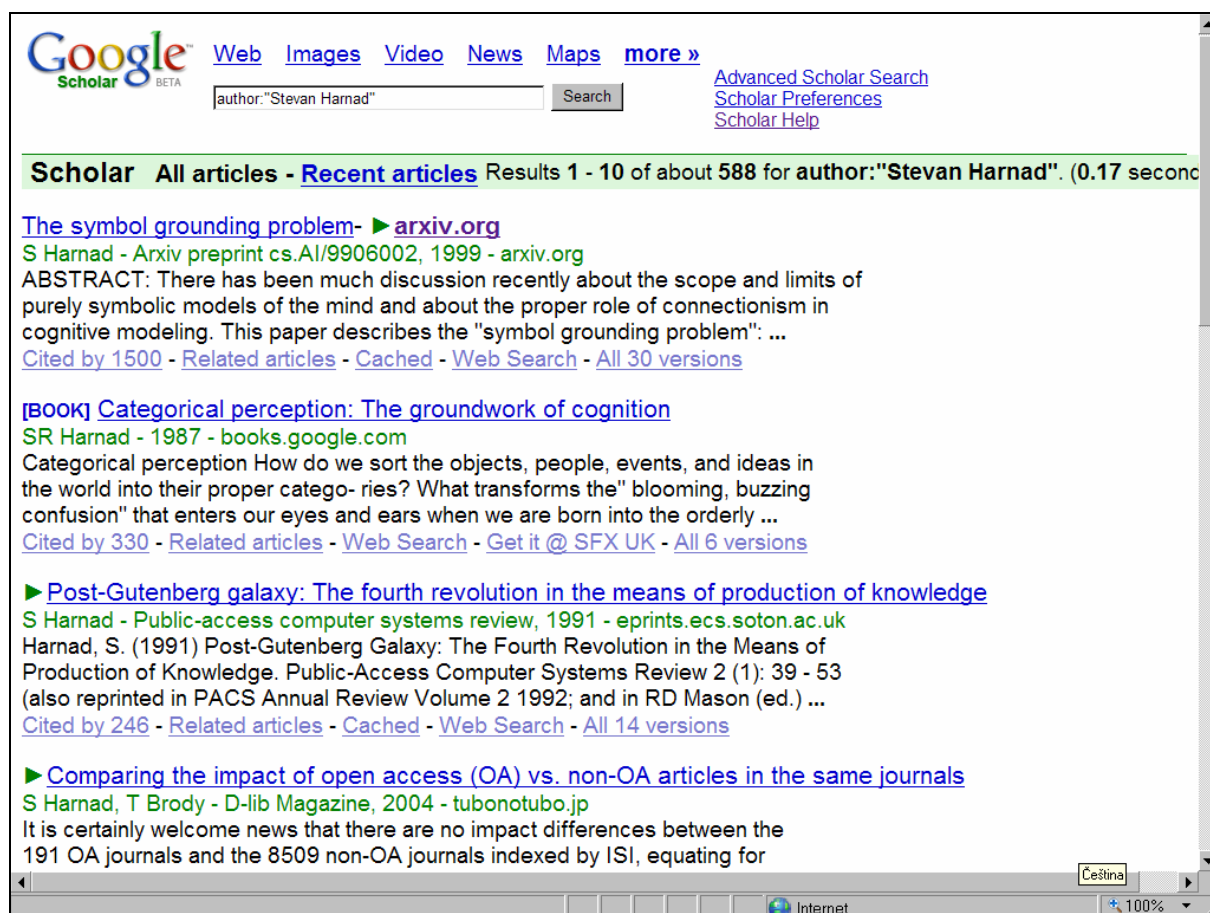
4.6.3.1 Google Scholar

Společnost **Google** (<http://www.google.com>) spustila v listopadu 2004 do provozu nový systém pro zpřístupňování vědecké literatury **Google Scholar** (<http://scholar.google.com/>). Systém je bezplatný a zahrnuje také **citační služby**. Jeho uvedení bylo doprovázeno nadšením i nedůvěrou a četnými otázkami, které se týkaly rozsahu a pokrytí jeho databáze. I po více než roce a půl však řada těchto otevřených otázek zůstává nezodpovězena [NEUHAUS, 2006, s 127]. Volně přístupná databáze systému Google Scholar je navíc pro světové uživatele k dispozici stále pouze v beta verzi (po čtyřech letech provozu). Nicméně popularita a návštěvnost systému je velká.

Systém Google Scholar je výsledkem úsilí skupiny vydavatelských společností, jež zajistily **obsah** (především metadata, tj. dodali určité části ze svých databází bibliografických záznamů), a společnosti Google, která se postarala o **programové a technické zabezpečení** systému. Společnost Google ovšem poskytuje zatím, na rozdíl od největších komerčních citačních služeb, jen **velmi málo informací** o svém novém produktu. Není zcela jasné obsahové zaměření databáze, její tvorba, způsoby doplňování, softwarové zabezpečení aj.

Celkový **počet zdrojových záznamů** v databázi Google Scholar není znám (jde zřejmě řádově o miliony záznamů). Znáno není ani celkový počet zastoupených vydavatelů, preprintových či reprintových serverů nebo samotných databází. Mnoho odborníků se pokouší různými způsoby zjistit tuto informaci. V jedné ze **studií z roku 2006** [NEUHAUS, 2006], která prezentovala výsledky měření rozsahu a obsahu databáze (porovnávána byla s obsahem 47 jiných vědeckých databází), její autoři zjistili, že Google Scholar má ve své beta verzi

některé silné stránky, ale také řadu stránek slabých. Pokrytí online časopisů s otevřeným přístupem, zdarma zpřístupňovaných vědeckých databází a jednotlivých vydavatelských databází je poměrně veliké. Zastoupení databází z humanitních disciplín a umění je však minimální, zastoupení databází ze společenských věd, výchovy a obchodu je zhruba 50% a je spíše nahodilé. Největší množství literatury je zastoupeno z oblasti přírodních a aplikovaných věd a medicíny (převaha volně dostupných zdrojů z daných oborů je zjevná). Slabé je rovněž pokrytí ne-anglicky psané literatury a chybí starší literatura. Při rešerších tedy uživatelé mnohdy netuší, že nalézají jen určitou část z celkového množství existující literatury. V databázi se vyskytuje také duplicita řady záznamů.



Obr. č. 4/37: rešerše prací vysoce citovaného autora Stevana Harnada a také citací jeho prací v systému Google Scholar (beta verzi); záznamy zahrnují hyperlinky na záznamy citujících autorů [2009-03-25]

Google Scholar nabízí především **vyhledávání záznamů zdrojových dokumentů** s možností návazného získání **úplných textů** z řady volně dostupných online zdrojů. Mnohé volně dostupné dokumenty však stále nejsou registrovány, stejně tak chybí řada vazeb na existující úplné texty. Systém proto intenzivně spolupracuje s mnoha knihovnami a dalšími dodavatelskými službami na **propojování záznamů s fondy úplných textů**, které tyto subjekty vlastní (viz například hyperlink u druhého záznamu na obr. č. 4/37 k propojení na rozhraní Jednotné informační brány Univerzity Karlovy v Praze - „Get it@SFX UK“). Využito je technologií SFX a OpenURL.

Systém standardně nabízí rozhraní pro **jednoduché a pokročilé vyhledávání**. Vyhledávat lze podle libovolného klíčového slova, autora zdrojového článku, názvu zdrojové publikace, data vydání a dokonce podle široké předmětové kategorie (nabízeno je 7 oblastí poznání).

V rámci vyhledávacího procesu lze předem nastavit další kritéria: jazyk komunikace se systémem Google (aktuálně 13 nabízených jazyků), jazyk dokumentu, nastavení propojení na fondy určité knihovny, počet výsledných záznamů zobrazovaných na stránce a nastavení linku k exportu záznamu do bibliografických manažerských systémů (aktuálně 5 nabízených systémů). Řazení nalezených záznamů je dáno specifickým algoritmem systému Google („Page Rank“), který je postaven především na webové citovanosti dokumentu, a do výpočtu je navíc zahrnuta také jeho klasická citovanost. Jiná pořádání záznamů, kromě možnosti zobrazení nejnovějších dokumentů (Recent articles), zatím nejsou nabízena.

Jak ve své analýze systému Google Scholar z listopadu 2005 uvedl P. Jacsó, systém Google Scholar ve své beta verzi vykazuje **při vyhledávání řadu problémů**. Některé rešeršní funkce jsou chybně koncipované a špatně implementované a výstupy záznamů obsahují mylné informace [JACSÓ, 2005a, s. 1538]. Problémy byly vykázány při využívání základní booleovské operace OR. Značné nejasnosti se týkají také vyhledávání záznamů podle roku vydání či vytvoření dokumentu (systém například při postupném zvětšování časového období předkládá postupně menší počty výsledných záznamů). Testovací rešerše všech dokumentů za období let 868 až 2005, provedená P. Jacsó v roce 2005, vedla k číslu zhruba 1 360 000 záznamů. Zdá se, že šlo o mylné číslo celkového počtu záznamů (ve srovnání se systémem WOS a Scopus by šlo o velmi malý počet). Zavádějící výsledky byly zjištěny rovněž u citovanosti dokumentů [JACSÓ, 2005a, s. 1541].

Google Scholar nabízí zdarma službu sledování **citovanosti zdrojového dokumentu**, která je reprezentována hyperlinkem „**Cited by**“ (viz obr. č. 4/37) vedoucím k záznamům citujících dokumentů. Je potřeba ale brát do úvahy fakt, že jde o citovanost měřenou pouze na základě dokumentů, které jsou indexovány systémem Google Scholar. Specifické záznamy začínající návěštím „**[Citation]**“ reprezentují záznamy citované literatury, které byly extrahovány ze seznamů použité literatury zdrojových dokumentů (nejsou prozatím předmětem indexace). Záznamy těchto citací ale nejsou klikatelné na primární dokumenty přímo, je nutné použít hyperlinku „**Web Search**“ a titul dohledat na Internetu. V databázi se vyskytují také duplicity citací. Celkový počet citací není veřejnosti znám.

Systém zatím také nenabízí žádné možnosti bibliometrických analýz databáze Google Scholar. Nelze zjistit její velikost a složení, tak jak tomu je u systému WOS a Scopus. K dispozici není ani seznam vydavatelů, ani seznam zdrojových časopisů nebo sborníků. Není udána retrospektiva databáze ani pokrytí zastoupených oborů.

Systém Google míří k vytvoření velkého zdroje vědecké literatury. Očekává se však podstatné doplnění informačních zdrojů (včetně materiálů z elektronických volně dostupných archivů), zlepšení vyhledávacího procesu systému (včetně vytvoření specifických rejstříků pro řadu údajů zahrnutých bibliografických záznamů). Rozšířeny by měly být možnosti řazení záznamů. Konsolidovány by měly být také citace na základě rozrůstajícího se registru DOI (o projektu zvaném CrossRef Search viz [BRATKOVÁ, 2006b]). Měly by být zavedeny nástroje pro zpracování záznamů citujících a citovaných dokumentů.

4.6.3.2 CiteSeer

Zajímavý směr v oblasti zpracování a využívání citací v prostředí digitálních knihoven je zastoupen systémem **CiteSeer.IST** (<http://citeseer.ist.psu.edu/>) a jeho novou verzí (beta) **CiteSeerX** (<http://citeseerx.ist.psu.edu/>). Systém dvakrát změnil jméno (původně se nazýval CiteSeer, později také ResearchIndex). Byl vytvořen v roce 1997 tehdejšími zaměstnanci

Výzkumného ústavu společnosti NEC (Princeton, N.J., USA) – S. Lawrencem, L. Gilesem a K. Bollackerem [LAWRENCE, 1999a]. V současnosti je systém již provozován Školou informačních věd a technologií (Pensylvánská státní univerzita) a jeho správou a dalším rozvojem je pověřen L. Giles (na jaře 2005 byly získány grantové prostředky ve výši 1,2 milionů dolarů na obohacení a zlepšení funkčnosti celého systému [JACSÓ, 2005b]; nová, zatím beta verze systému byla spuštěna v roce 2008).

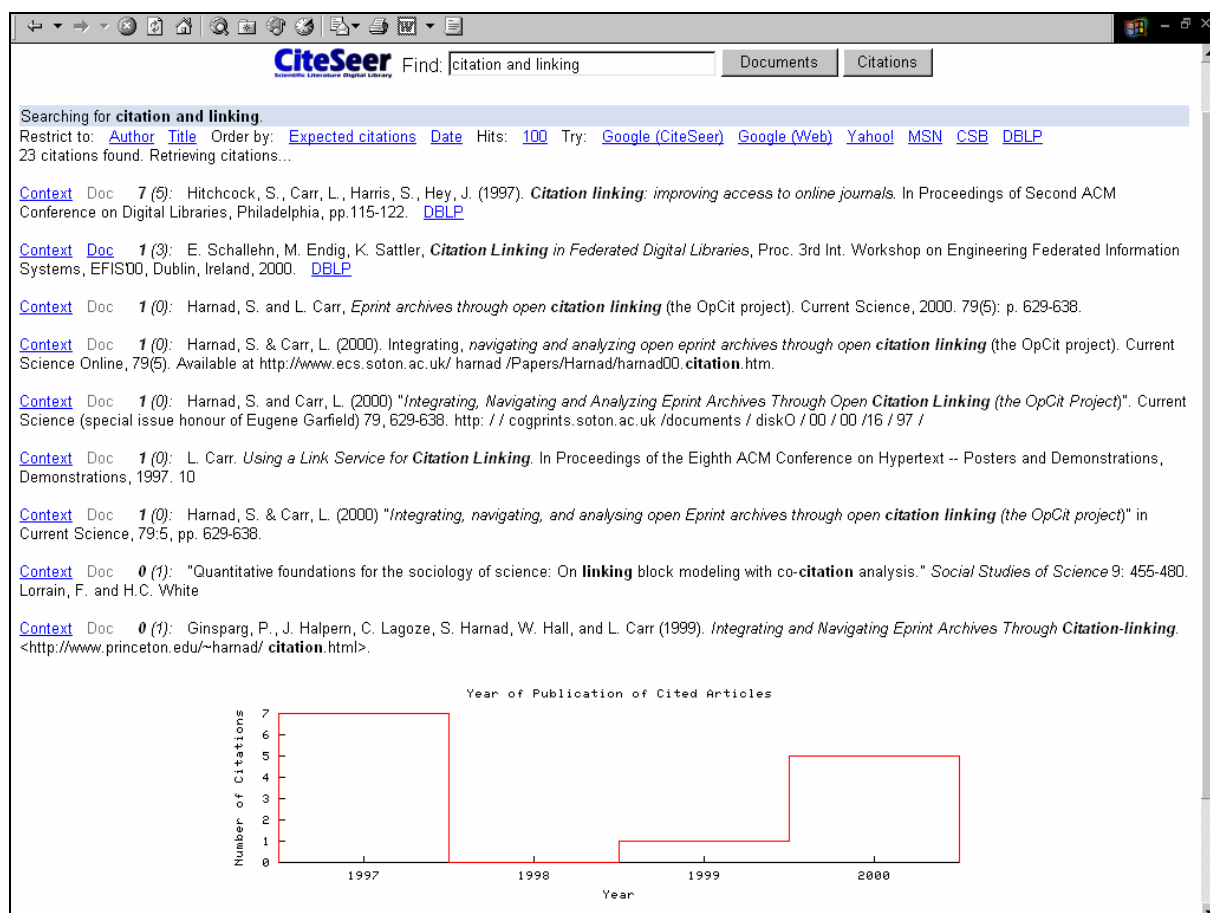
Specifikem systému je **citační rejstřík** budovaný na bázi digitální knihovny, která zahrnuje volně dostupné dokumenty v elektronické formě z oboru **počítačové vědy** a nově také z oboru **informační vědy** v celosvětovém záběru. Získávanými zdrojovými dokumenty jsou volně dostupné preprinty statí z konferencí nebo časopiseckých článků, výzkumné zprávy, disertace aj. Limitem jsou formáty získávaných typů dokumentů (PostScript a PDF). V oboru počítačové vědy jde ale o podstatné formáty, které pomáhají udržovat vysokou kvalitu databáze. Systém sice zahrnuje vlastního crawlera pro zjišťování a stahování dokumentů z WWW, zatím jej však nevyužívá. V systému se praktikuje jejich rychlejší vyhledávání pomocí známých vyhledávačů (Alta Vista, HotBot, Excite aj.). Využívá se strategie volby kombinace účinných selekčních údajů (jde např. o anglická slova „publications“, „conferences“, „proceedings“, „papers“, „postscript“ aj.). Monitorovány jsou pravidelně také elektronické diskusní skupiny. Využívá se i přímého nahlašování dokumentů k indexaci. Specifickou funkcí systému je automatické zjišťování a vyřazování **duplicitních dokumentů**.

Stažené dokumenty jsou konvertovány do ASCII formátu pomocí programu PreScript, který také používá známá novozélandská digitální knihovna NZDL. Automaticky se ověřuje, zda dokument má či nemá vědeckou či výzkumnou povahu (testuje se přítomnost citací nebo jejich sekcí). Významnou etapu ve zpracování zdrojových dokumentů představuje jejich indexace a tvorba běžného invertovaného souboru pro potřeby vyhledávání (vedle standardních se využívá i některých specifických technik s ohledem na funkce systému). Systém nevyužívá slovník stop slov z důvodu zvýšení přesnosti při vyhledávání (u autorů je např. často nutné v dotazu uvést iniciály jejich křestního jména).

Veškeré procesy v systému probíhající jsou zajišťovány programovými prostředky. Unikátní technologie zajišťuje **autonomní indexaci citací (ACI, Autonomous Citation Indexing)**. Program ACI zajišťuje v systému tvorbu citačního indexu, který se v základních parametrech podobá komerčnímu citačnímu rejstříku SCI. Ve zdrojových dokumentech (oproti SCI jsou zahrnuty nejen články z časopisů) se zcela automaticky (nikoliv manuálně) na základě analýzy textu zjišťují seznamy bibliografických odkazů (vodítkem může být identifikace hlavičky seznamu nebo seznam samotný). Následuje **extrakce jednotlivých citací**. Jsou identifikovány buď na základě jednoduchých **citačních identifikátorů** nebo jenom na základě **vertikálního mezerování** či **odsazování** záznamů. Extrahovány jsou údaje jako autor a název citovaného dokumentu, název zdrojového dokumentu, rok publikování, čísla stránek a zejména již zmíněný **citační identifikátor** (označení typu „[12]“, „[Boll99]“, „(Cameron 1997)“ aj.), pomocí něhož lze poté lokalizovat citaci v textu. Na základě citace je možné v textu zdrojového dokumentu automaticky extrahovat i její kontext, tj. množinu slov v okolí citačního identifikátoru. Systém je schopen zpracovat také různé varianty citačního identifikátoru v případě, že v citaci jsou uvedeni všichni autoři nebo jenom první z nich. Využívá se pravidelně se opakujících výrazů v citaci. K identifikaci údajů citací se využívá i doplňkové databáze jmen autorů, názvů časopisů apod.

Bibliografické odkazy na stejný dokument mohou být různě strukturovány a mohou zahrnovat různá množství údajů a také různá uspořádání těchto údajů. Význam programu ACI spočívá především ve schopnosti rozpoznat, že citace odkazují na stejný dokument. Využito je heuristického postupu založeného na **automatickém rozkladu citace**. Stálé výrazy, tj. údaje, které mají relativně stejnou syntax, pozici a složení, se zpracovávají jako první. Může se např.

zjistit, že citační identifikátor vždy předchází celému záznamu a že zůstává nezměněn ve všech ostatních záznamech. Jakmile program zjistí pravidelné vlastnosti u nějaké citace, využije trendů v syntaktických vazbách mezi údaji pro předvídání místa, kde požadovaný údaj existuje, existuje-li vůbec. Např. informace o autorovi většinou vždy předchází názvu citované práce apod. Na základě této schopnosti je systém CiteSeer schopen generovat seznamy citací (citovaných prací viz obr. 4/38) a také statistiky citační frekvence [LAWRENCE, 1999b].



Obr. č. 4/38: přehled záznamů citací včetně grafické prezentace na téma „citace a propojování“ v systému CiteSeer – v uspořádání podle citovanosti daných dokumentů [získáno 2006-04-28]

Na základě technologie ACI je tedy vytvářen specifický index citací (vedle základního indexu úplných textů zdrojových dokumentů). Program ACI umožňuje zjišťovat a extrahovat z dokumentů citace, identifikovat citace reprezentující stejný dokument, byť se vyskytují v různých formách, a také identifikovat zajímavé **kontexty citací** ve zdrojových dokumentech. Dokumentace systému proklamuje zjištěnou **chybovost** jenom cca 5 %. (Podrobnosti o zpracování informací v systému CiteSeer viz také publikovaný článek autorky z roku 2001 [BRATKOVÁ, 2001]). Databáze zahrnovala v době před otevřením nové verze CiteSeerX **767 558** úplných textů zdrojových dokumentů a řádově **miliony citací** (počet nebyl na WWW stránkách systému uveden). Uživatelům systém poskytuje zajímavé výstupy a služby, řada z nich je zcela nového charakteru.

Systém CiteSeer.IST nabízel v původní verzi několik málo způsobů vyhledávání informací. V rámci přímé formulace dotazů může být uplatněno libovolné klíčové slovo, dotaz lze uplatnit jak vůči indexu zdrojových dokumentů, tak indexu citací. Zajímavé výstupy nabízí

vyhledávání citací (viz obr. č. 4/38). Výsledkem je rešerše **sjednocených záznamů citovaných dokumentů** (doplněná grafickou prezentací vývoje citovanosti na dané téma), ze kterých se lze dostat pomocí specifické funkce (Context) ke zkráceným záznamům zdrojových dokumentů (zahrnují části textů lokalizovaných okolo citačního identifikátoru a původní záznam citace). Ze zkráceného záznamu se lze pomocí hyperlinku dostat k úplnému záznamu zdrojového dokumentu a z něho k úplnému textu (v různých formátech). Seznamy záznamů citací lze třídit na základě mnoha údajů, k nejzajímavějším patří uspořádání podle **očekávané citovanosti** (Expected citations - viz obr. č. 4/38). Systém nabízel také několik zajímavých statistik: seznam nejvíce využívaných zdrojových dokumentů, nejvíce citovaných zdrojových nebo nejvíce citovaných dokumentů (vlastní citační rejstřík), seznam nejvíce citovaných autorů aj. Zajímavostí je také automaticky vytvářený předmětový rejstřík (zatím jen pro počítačovou vědu). K návazným funkcím systému CiteSeer.IST patří služba **průběžného informování** uživatelů o nových přírůstcích ve zdrojových dokumentech nebo citacích.

Systém CiteSeer představuje zajímavý experiment v současném rozvoji citačních služeb na Internetu. Reprezentuje modelovou situaci výrazného trendu propojování informací a integrace informačních systémů. Přispívá k celkovému zlepšení organizace, zpracování, vyhledávání, rozšiřování a zpřístupňování vědecké literatury v rámci sítě Internet. Kvalita softwaru byla oceněna také tím, že společnost Thomson Reuters si ho vybrala jako základ pro svůj nový projekt „Web Citation Index“.

Indexing and retrieval of scientific literature (1999) [60 citations — 14 self]
 by Steve Lawrence, Kurt Bollacker, C. Lee Giles
 In Eighth International Conference on Information and Knowledge Management, CIKM 99
<http://www.neci.nj.nec.com/~lawrence/papers/cs-cikm99/cs-cikm99.ps.gz>
 External Links: DBLP
 Add To MetaCart

Abstract:
 The web has greatly improved access to scientific literature. However, scientific articles on the web are largely disorganized, with research articles being spread across archive sites, institution sites, journal sites, and researcher homepages. No index covers all of the available literature, and the major web search engines typically do not index the content of Postscript/PDF documents at all. This paper discusses the creation of digital libraries of scientific literature on the web, including the efficient location of articles, full-text indexing of the articles, autonomous citation indexing, information extraction, display of query-sensitive summaries and citation context, hubs and authorities computation, similar document detection, user profiling, distributed error correction, graph analysis, and detection of overlapping documents. The software for the system is available at no cost for non-commercial use. 1

Citations

- 1835 [The anatomy of a large-scale hypertextual Web search engine](#) - Brin, Page - 1998
- 1687 [Authoritative sources in a hyperlinked environment](#) - Kleinberg - 1998
- 1043 [The PageRank Citation Ranking: Bringing Order to the Web](#) - Page, Brin, et al. - 1998
- 608 [Managing Gigabytes: Compressing and Indexing Documents and Images](#) - Witten, Moffat, et al. - 1999
- 363 [Improved algorithms for topic distillation in a hyperlinked environment](#) - Bharat, Henzinger - 1998
- 266 [Syntactic clustering of the web](#) - BRODER, GLASSMAN, et al. - 1997
- 220 [Efficient crawling through URL ordering](#) - Cho, Garcia-Molina, et al. - 1998
- 177 [Multi-service search and comparison using the MetaCrawler](#) - Selberg, Etzioni - 1995
- 172 [Accessibility of information on the web](#) - Lawrence, Giles - 1999
- 140 [CiteSeer: an automatic citation indexing system](#) - Giles, Bollacker, et al. - 1998
- 117 [Learning hidden markov model structure for information extraction](#) - Seymore, McCallum, et al. - 1999
- 116 [Copy detection mechanisms for digital documents](#) - Brin, Davis, et al. - 1995

POPULAR TAGS
 Add a tag:
 No tags have been applied to this document.

BIBTEX | ADD TO METACART

```
@INPROCEEDINGS{Lawrence99indexingand,
  author = {Steve Lawrence and Kurt Bollacker and C. Lee Giles},
  title = {Indexing and retrieval of scientific literature},
  booktitle = {In Eighth International Conference on Information and Knowledge Management, CIKM 99},
  year = {1999},
  pages = {139--146}}
```

YEARS OF CITING ARTICLES

Year	Citations
2000	5
2001	2
2002	3
2003	2
2004	3
2005	1

Obr. č. 4/39: část úplného záznamu jednoho z článků tvůrců systému CiteSeer ve výstupním prezentačním formátu nové verze systému CiteSeerX [2009-03-25]

Nová verze systému CiteSeer zvaná **CiteSeerX** (<http://citeseerx.ist.psu.edu/>) byla uživatelům zpřístupněná v roce 2008 (zatím v beta verzi). Nové přírůstky systému jsou již zpracovávány jen do nové báze. Ke konci března 2009 systém zahrnoval již 1 366 867 zdrojových dokumentů a 26 435 805 citací. Systém má novou koncepci zpracování záznamů (na bázi XML), metadata systém nabízí i dalším informačním službám volně přes rozhraní OAI-PMH (bázová adresa: <http://citeseerx.ist.psu.edu/citeseerx/oai2>). Nová je koncepce výstupních služeb pro koncové uživatele. Připravena jsou již dvě rozhraní (základní a pokročilé), záznamy zdrojových dokumentů jsou zobrazovány v novém designu (viz obr. č. 4/39). Záznamy citací se vyhledávají nyní najednou se záznamy zdrojových dokumentů (nikoliv zvlášť). Úplné texty jsou nabízeny v různých formátech přímo z digitálního skladu systému CiteSeerX.

4.6.3.3 Citebase Search

Z mnoha hledisek pozoruhodná citační služba **Citebase Search** (<http://www.citebase.org/>) je produktem původního americko-britského projektu **OpCit** (The Open Citation Project, <http://opcit.eprints.org/>). Jak ve svém hodnocení uvedl P. Jacsó, Citebase Search je jeho „královským klenotem“ [JACSÓ, 2004a, s. 57]. Projekt OpCit (1999-2002) se zabýval otázkami vnitřního a vnějšího propojování bibliografických záznamů a citací uložených ve volně dostupných archivech elektronických tisků. Jeho neméně důležitým cílem bylo také zajišťování komplexních bibliometrických či informetrických analýz, které pomáhají lépe porozumět tomu, jak je vědecká literatura využívána v tvůrčím procesu.

Systém **Citebase Search** byl vyvinut na Southamptonské univerzitě (Velká Británie), jeho tvůrce a správce je Tim Brody. Jeho existence byla odborné veřejnosti poprvé oznámena v prosinci 2001, uživatelé jej však mohli začít využívat až po jeho integraci s archivem arXiv.org v srpnu 2002. Citebase Search je autonomním scientometrickým nástrojem ke zjišťování potenciálu informačních materiálů uložených v elektronických archivech s otevřeným přístupem. Systém je stále v **experimentální fázi** svého vývoje, jak je uvedeno na jeho hlavní webové stránce, nelze ho zatím využívat například pro potřeby reálných akademických evaluací, protože pokrytí citací a jejich analýza není (prozatím) úplná. Jde o jeden z prvních systémů, který začal pracovat na bázi principů OAI (jako poskytovatel služby).

Do databáze Citebase Search jsou pomocí protokolu OAI-PMH sklízena **metadata zdrojových dokumentů** z několika elektronických archivů. Metadata jsou získávána ve formátu Dublin Core a běžně k nim patří název dokumentu, abstrakt nebo komentář, jména autorů (v nestrukturovaném tvaru, příjmení jsou ale zpravidla odlišitelná pro potřeby dalšího zpracování), lokace zdroje (např. URL), data různých typů (datum revize aj.). V některých archivech jsou součástí sklizených metadat také **citace** ve formě volného textu (jsou „strukturované“ pouze autory podle různých citačních stylů). Název a abstrakt zdrojového dokumentu vcházejí do invertovaného souboru pro potřeby volného vyhledávání v úplném textu. Systém Citebase Search však také na základě **vlastního rozhraní** sklízí úplné texty zdrojových dokumentů (nejsou ale v systému natrvalo ukládány), ze kterých **extrahuje záznamy citovaných dokumentů**, jež jsou v rámci dalšího zpracování s těmito dokumenty propojovány, jsou-li volně na WWW lokalizovány [BRODY, 2003, s. 4]. Extrahovány jsou běžně přesné názvy časopisů (často zkrácené), rok publikování, ročník, číslo a stránkování. Extrakce citací probíhá rozdílně podle toho, v jakém formátu jsou k dispozici úplné texty. Snadno lze například získat strukturované citace z archivu BioMed Central (dokumenty jsou ve formátu XML). Složitější způsob získání citací je nasazen například u většiny textů z archivu arXiv.org (texty jsou ve zdrojovém formátu LaTeX, popř. PDF, nutná je tedy jejich konverze do jednoduchého textu před vlastní extrakcí citací). K rozpoznávání záznamů citací

a jejich dílčích údajů je využíváno specifických algoritmů. Sklizená metadata a extrahované záznamy citací slouží k tvorbě citační databáze, která je základem následných webových služeb včetně možného exportu dat pomocí protokolu OAI-PMH a tvorby datových analýz [BRODY, 2003, s. 5].

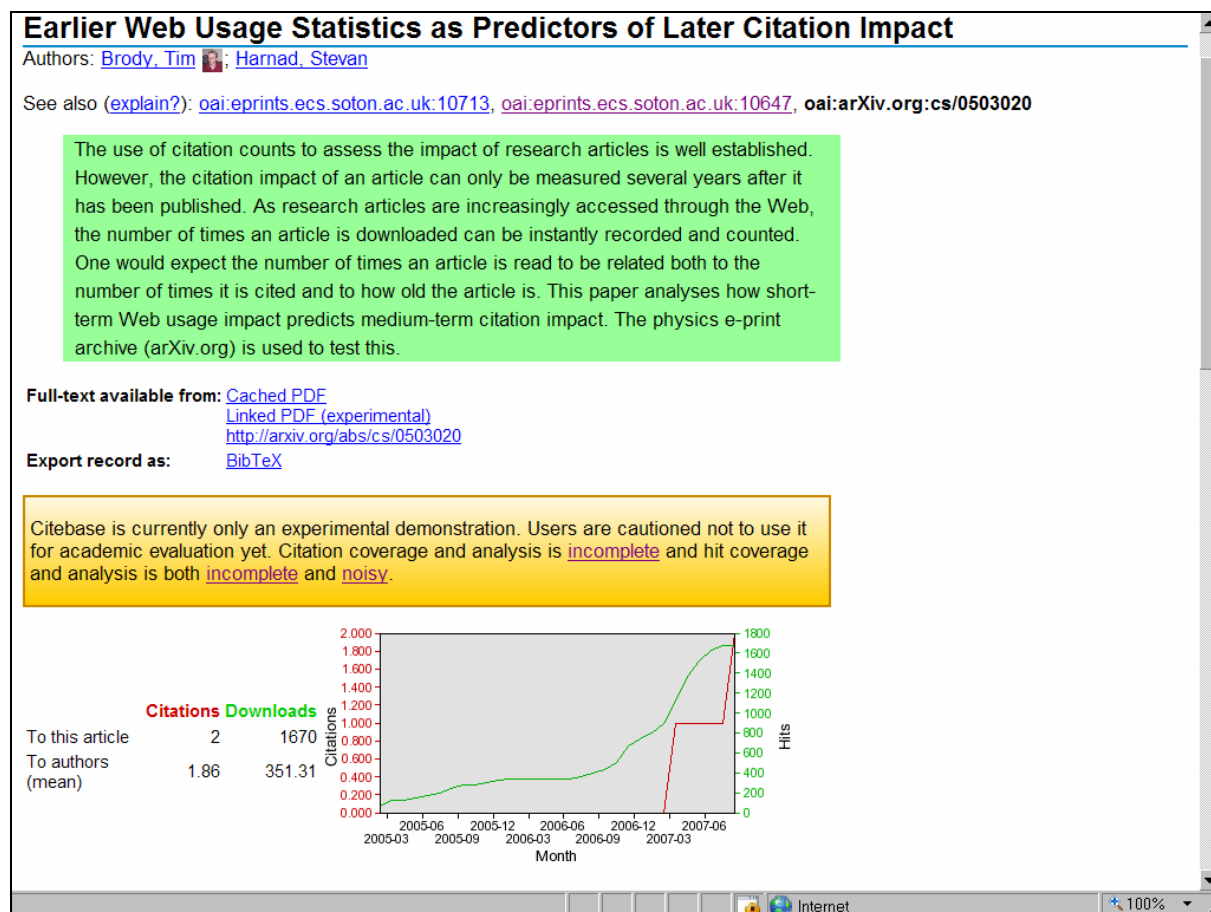
V dubnu 2006 systém Citebase Search ohlásil téměř **465 000 záznamů zdrojových dokumentů** (s propojením na jejich úplný text) – aktuální stav není na WWW k dispozici. Největší počet pochází z partnerského archivu **arXiv.org**, který byl původně jediným archivem projektu. Další záznamy pocházejí například z archivu **ECS Southamptonské univerzity** a také archivu **CogPrints Sotonské univerzity**. Sklizeny byly také archivy systému PubMed Central a BioMed Central, archiv E-LIS pro knihovní a informační vědu - v současné chvíli je jejich sklizení pozastaveno. Předpokládá se, že v budoucnu by systém Citebase Search mohl pokrýt všechny volně dostupné elektronické archivy. Počet **registrovaných citací** byl v dubnu 2006 téměř **13 380 000**, více než **3 000 000** z nich bylo propojeno na úplný text. V rámci citací je identifikováno (příjmení, křestní jméno) zhruba 350 000 autorů. Ve stejné době systém zaregistroval již **5 640 129** stažení úplných textů.

The screenshot shows the Citebase Search web interface. At the top is a navigation bar with links: Search Citebase, Information and Help, Impact Health Warning, Login/Register, and a search box. Below this is the 'Search Results' section. It features three tabs: Metadata, Citation, and Identifier. The 'Metadata' tab is active, showing search criteria: Authors' name(s) as 'Brody Tim', Title or Abstract Keywords as 'Web', and Publication Title as an empty field. There are also fields for 'Record Year' and 'Rank matches by' (set to Descending). A 'Search' button and a 'Reset' button are at the bottom of the criteria section. Below the search criteria, a status bar indicates 'Showing 1 - 10 of 11 found' and 'Query took 0.357 seconds'. The results list shows three entries, each with a title, citation information, and a brief description. The first entry is 'Open Citation Linking: The Way Forward' by Hitchcock, Steve; Brody, Tim; Gutteridge, Christopher et al. The second is 'A Scalable Architecture for Harvest-Based Digital Libraries - The ODU/Southampton Experiments' by Liu, Xiaoming; Brody, Tim; Harnad, Stevan et al. The third is 'Earlier Web Usage Statistics as Predictors of Later Citation Impact' by Brody, Tim; Harnad, Stevan; Carr, Les. The interface is displayed in a web browser window with a status bar at the bottom showing 'Internet' and '100%' zoom.

Obr. č. 4/40: zkrácené záznamy dokumentů (autora Tima Brodyho) uložené v databázi Citebase Search, uspořádané podle jejich citovanosti [2009-03-25]

Rozhraní pro vyhledávání informací systému Citebase Search nabízí standardně **vyhledávání v indexu metadat**, a to podle jména autora, klíčových slov z názvu/abstraktu zdrojového článku, názvu časopisu a data vytvoření nebo publikování dokumentu. Zvláště lze vyhledávat podle citací a podle OAI identifikátoru (viz například oai:arXiv.org:cs/0503020 v záznamu na obr. č. 4/41). Výsledky rešerší vedou vždy k seznamu zkrácených záznamů zdrojových

dokumentů (obr. č. 4/40). Specifickou nabídkou systému je **uspořádání či řazení nalezených záznamů** (Rank matches by - viz také obr. č. 4/40): podle citačního vlivu článku (Citations (Paper)) (obr. č. 4/40), citačního vlivu autorů článku (Citation (Author)), počtu webového stažení článku (Hits (Paper)), počtu webového stažení dokumentů autorů článku (Hits (Author)), data vytvoření článku (Date (Creation)) a data poslední aktualizace (Date (Update)).



Obr. č. 4/41: část obrazovky s úplným záznamem dokumentu z databáze Citebase Search včetně údajů a grafu o vývoji citovanosti vyhledaného titulu [2009-03-25]

Ze zkráceného záznamu zdrojového dokumentu lze pomocí hyperlinku získat **úplný záznam**, který se vyznačuje bohatostí poskytovaných informací (viz obr. č. 4/41). Zahrnuje úplné bibliografické údaje, zpravidla vždy abstrakt, URL s uloženým úplným textem a kompletní citační informace. V přehledné tabulkové formě, která je generována korelačním generátorem, je uveden počet citací a webových stažení daného článku a průměrný počet citací a průměrný počet webových stažení všech dokumentů autorů daného článku. Grafická informace zobrazuje uvedené parametry v čase. V dalších sekcích úplného záznamu jsou k dispozici seznamy citací z daného článku, seznamy citací na daný článek a seznamy kocitací článků.

Systém Citebase Search nabízí uživatelům sofistikované rešeršní prostředí, poskytuje specifické bohaté citační služby týkající se volně dostupné vědecké literatury. Data z databáze lze využít k dalším četným scientometrickým, bibliometrickým a informetrickým analýzám, byť jde zatím o experimentální provoz. V budoucí reálné verzi bude potřebné rozšířit obsah databáze o informace z dalších archivů.

4.6.4 Závěr

Vybrané volně dostupné citační systémy, analyzované výše v textu, poskytují koncovým uživatelům užitečné a zajímavé **výstupy a služby**. Nejnovější systémy mají moderně řešená rozhraní pro vyhledávání a nové nástroje pro navigaci z výsledků rešerší. Některé systémy nabízejí webové linky ze záznamů citované literatury (Citebase Search). U některých systémů zatím chybí nástroje pro analýzu výsledných rešerší (Google Scholar). Efektivní seznamy kcitovaných dokumentů (tj. dokumentů, které byly společně citovány aktuálním a nějakým jiným dokumentem) zajišťuje služba Citebase Search. Obsahově bohatý fond úplných textů má systém CiteSeer, může proto nabízet velkou škálu výstupů a návazných bibliometrických analýz. Rozbor systémů zároveň ukazuje, že v současnosti dochází k velké **multiplicitě při získávání a propojování citací** stejných titulů literatury včetně literatury volně dostupné na WWW (týká se i komerčních služeb). Citační systémy poskytují podobné, ale zároveň limitované služby. Není jasné, jak by vlastně měl být budován **optimální citační rejstřík pro volně dostupnou literaturu**. Systémy Citebase Search a CiteSeer jsou založeny téměř výlučně na elektronických tiscích archivovaných samotnými autory. Záznamy citované literatury jsou v tomto případě k dispozici v **nestrukturované formě**, a proto musejí tyto citační systémy zavádět určité algoritmy k rozkladu výchozích citačních textů a k vytvoření jednotné formy záznamu citovaného dokumentu, která bude vhodná k automatickému propojování na dokumenty nebo jiné záznamy. U systému Google Scholar není postup získávání a propojování citací veřejnosti znám, z provozu systému lze však usuzovat, že má k dispozici také vlastní algoritmus k rozkladu nestrukturovaných citací z úplných textů a jejich propojování.

Nárůst volně dostupných vědeckých dokumentů ukládaných do stále se rozrůstajícího velkého počtu centralizovaných elektronických archivů nebo institucionálních repozitářů pokračuje v globálním rámci rychlým tempem. Význam této literatury je značný až do té míry, že se staly také předmětem zpracování významných komerčních citačních služeb WOS a Scopus. Jde však o služby drahé, dostupné jen na základě předplatného, které navíc zpracovávají jen vybranou množinu těchto volně dostupných informačních materiálů. Je proto potěšitelné, že vznikly a úspěšně se rozvíjejí nové, moderně koncipované volně dostupné citační služby, které těží z toho, že zpracovávány materiály včetně metadat jsou v jen digitální formě, že jsou uloženy v dobře organizovaných repozitářích (digitálních skladech) a že jsou volně dostupné. Zahrnutí úplných textů do zpracování v rámci těchto citačních služeb poskytuje, jak ukazuje analýza, mnohem větší možnosti při výstupním zpracování všech uložených informací. Konečnou cílovou službou jsou scientometrické, bibliometrické a informetrické analýzy, které budou pomáhat odhalovat vliv volně dostupných materiálů na vývoj vědy.

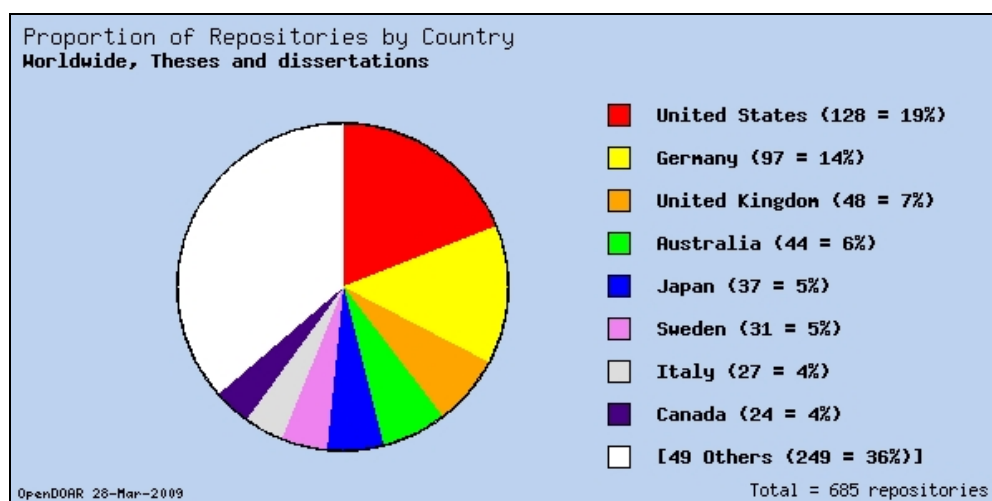
Hlavním problémem, kterým je nutné se v současnosti zabývat, je optimalizace tvorby záznamů citací, jejich struktura pro automatické strojové zpracování a automatické propojování na samotné dokumenty. V současné době se záznamy citací získávají v systémech citačních služeb především pomocí specifických algoritmů metodou extrakce a rozkladu přímo z úplných textů, což není optimální (mohou nastat problémy s identifikací a chybovost může celkově velká). Jde o různorodé ručně zapisované záznamy v podobě nestrukturovaných textů a je úkolem programu, aby dokázal automaticky sjednotit různé podoby citací na stejný dokument. Je tedy nezbytné, aby se v dalším období odborníci zaměřili na problematiku standardizace citačních záznamů, které jsou ukládány v elektronických archivech. V rámci zdokonalování softwarů, které zajišťují chod archivů či repozitářů, by měly být nasazeny specifické moduly, které by v blízké budoucnosti umožnily automatickou extrakci a reformátování citací (dle stanoveného standardu) z textů bezprostředně po jejich uložení v procesu auto-archivace tak, aby je jejich autoři měli možnost ve strojovém formátu zkontrolovat a případně okamžitě opravit. Znamenalo by to

možnost jejich automatického sklizení a dalšího bezproblémového zpracování v databázích citačních rejstříků. Doporučovaným standardem pro strukturování citací je v současnosti standard OpenURL.

4.7 Digitální knihovny vysokoškolských kvalifikačních prací

4.7.1 Úvod

Tato podkapitola těží z víceleté práce autorky disertační práce v oblasti systémů pro zpřístupňování vysokoškolských kvalifikačních prací (dále také jako VŠKP) v elektronické formě. Již v roce 2002 připravila a zveřejnila v rámci doktorského studia rozsáhlejší článek o stavu zpřístupňování elektronických VŠKP v zahraničí [BRATKOVÁ, 2002a], v letech 2003-2004 pak rozsáhlejší analýzu (výzkumnou zprávu) ke stejnému tématu v rámci grantové úlohy FRVŠ č. 2045/2003 (okruh E2) „E-DI“ (Digitální knihovna diplomových a disertačních prací UK v Praze) [BRATKOVÁ, 2004a]. Text této zprávy byl zveřejněn i na serveru Odborné komise pro otázky elektronického zpřístupňování VŠKP (<http://www.evskp.cz/>) Asociace knihoven vysokých škol ČR (AKVŠ), kterou pomáhala založit (připravila také podklady pro důvodovou zprávu [Asociace knihoven, 2004a] a stanovisko AKVŠ k založení Odborné komise [Asociace, 2004b]) a již se stala odborným garantem. Činnost v Odborné komisi vyústila, kromě jiného, v roce 2008 také do práce na Centralizovaném projektu „Národní registr VŠKP a úložiště závěrečných prací se službou na odhalování plagiátů“, MŠMT C1/2008, konkrétně v oblasti přípravy národních metadatových standardů pro VŠKP (v pracovním týmu Knihovny Akademie múzických umění v Praze).



Obr. č.4/42: grafické znázornění podílu vybraných zemí na archivaci a zpřístupňování elektronických VŠKP [převzato z OpenDOAR, 2009-03-25]

Disertace a jiné typy VŠKP byly a stále jsou zařazovány ke specifické skupině literatury, již se v době jejich tradiční produkce dostalo označení „šedá literatura“ (angl. grey literature). Dřívější definice zdůrazňovaly její (tehdy významný) znak špatné dostupnosti. Národní a mezinárodní registrace této literatury v té době vznikaly proto, aby se zlepšila informovanost o jejich existenci a následně i možnost jejich získání. Doba však pokročila, technologie se změnily a rozšiřování informací i v oblasti vědy, výzkumu a vzdělávání zaznamenalo radikální proměny. Disertace jsou dnes vytvářeny jejich autory běžně v elektronické formě a řadu z nich je možné přes WWW najít, přečíst si a stáhnout zdarma (zejména v určitých zemích). Je tomu tak proto, že je jejich autoři dobrovolně nebo na základě nařízení školy uložili do volně dostupných digitálních archivů provozovaných na serverech univerzit. Zdůraznit je třeba ale fakt, že zatím jde jen o menší množství z celkového počtu všech disertací obhájených každoročně v celosvětovém rámci, nicméně jejich fonde se

každoročně zvětšuje. V rámci světové registrace digitálních repozitářů OpenDOAR (<http://www.opendoar.org/>) je v současné době (v březnu 2009) registrováno 685 repozitářů, které vykazují archivaci vysokoškolských kvalifikačních prací (disertačních, diplomových a popřípadě jiných prací, angl. Theses and Dissertations). Graf na obr. č. 4/42 ukazuje podíl jednotlivých zemí na archivaci a zpřístupňování elektronických VŠKP. „Lídry“ jsou Spojené státy americké, Německo, Austrálie a nově také Velká Británie (v loňském roce ještě nebyla v popředí žebříčku).

Samostatnou skupinu v oblasti digitálních knihoven však tvoří systémy, které obsahují jenom disertační a popř. i další vysokoškolské kvalifikační práce. Vytvářejí a rozvíjejí se jak na lokální úrovni jednotlivých univerzit, fakult nebo jejich ústavů, tak na úrovni regionální, národní nebo mezinárodní (v univerzálním nebo oborovém záběru). Pokud v tuto chvíli záměrně opomineme komerční větev zpřístupňování disertací v digitální formě (např. světově známou službu společnosti UMI z USA), pak volné zpřístupňování digitálních disertací, které se v posledních 10-12 letech intenzivně rozvíjí v různých částech světa, představuje zároveň i nový typ publikování těchto dokumentů, a to publikování elektronické. Je zřejmé, že volné zpřístupňování elektronických dokumentů má nesporné výhody, například při propojování systémů a budování národních metadatových registrů nebo jiných návazných nadstavbových služeb.

Je vhodné zdůraznit, že nově koncipované systémy pro rozšiřování elektronických disertací jsou systémy, jejichž budování a provozování zahrnuje nejen **knihovníky** a **informační pracovníky**, ale také **počítačové odborníky** z výpočetních center, **univerzitní učitele** a samozřejmě samotné **tvůrce** těchto prací. Má-li systém optimálně fungovat, je nutná souhra všech těchto aktérů [BRATKOVÁ, 2003a].

4.7.2 Zpřístupňování elektronických disertací v zahraničí

Následující část zahrnuje charakteristiku vybraných významných zahraničních systémů pro rozšiřování elektronických disertací a jiných VŠKP. Pro potřeby této práce byly vybrány dva mezinárodní systémy (služby), a to NDLTD s centrem v americké Virginii a systém CyberThèses, který má dvě centra, v Chile a ve Francii. A vybrány byly dále dva úspěšné národní systémy, a to australský systém ADT (bez systému povinného výtisku) a německý systém DissOnline.de (se systémem povinného výtisku). Jednotlivé dílčí části obsahují stručnou charakteristiku zvolených systémů a popis jejich vybraných parametrů. Na závěr je připojena informace o situaci v České republice.

4.7.2.1 Mezinárodní federace a systém NDLTD

Významným centrem pro zpřístupňování elektronických disertačních a diplomových prací je v současné době systém, provozovaný Virginským polytechnickým institutem a státní univerzitou (Virginia Polytechnic Institute and State University, <http://www.vt.edu>). Je znám jak pod zkráceným jménem **NDLTD**, tak pod jménem úplným „**Networked Digital Library of Theses and Dissertations**“ (Sít'ová digitální knihovna diplomových a disertačních prací). Zajímavé je, že v první fázi projektu v roce 1996 se objevilo nejprve jméno „National Digital Library of Theses and Dissertations“ (Národní digitální knihovna diplomových a disertačních prací) [FOX, 1996]. Předpokládalo se, že by mohl vzniknout národní systém USA pro disertační a diplomové práce (grantové prostředky poskytly i státní orgány a organizace). Systém ovšem měl a má k takovému označení daleko. Do dnešního dne se nepodařilo do systému zapojit řadu amerických univerzit. Představitelé systému NDLTD v roce 1996 uvedli, že v USA je každým rokem obhajováno cca 40 000 disertací a cca 360 000 diplomových prací [FOX, 1996, část 1, Introduction]. Z bilanční zprávy z roku 2001 lze vyčíst, že do digitálního

fondy NDLTD se za 5 let podařilo získat jenom cca 5 820 prací z USA (číslo nezahrnuje digitalizované starší práce a také práce ze zahraničí) [SULEMAN, 2001a, Collection Size]. Přestavuje to jenom malý zlomek z celkového fondu všech předložených prací za celé 5leté období.

Systému NDLTD, který vznikl z lokální aktivity Virginské techniky, se nicméně podařilo získat aktivní účast některých univerzit z USA i ze zahraničí, z pohledu jejich celkového počtu se ale zatím nedá označit v pravém slova smyslu jako světový či globální systém. Oficiálně se i na WWW stránkách NDLTD objevuje příznačnější označení systému: „federace“. To, co především přitahuje odborníky ze světa, je softwarová a technologická stránka systému. Byly připraveny nebo adaptovány a jsou využívány kvalitní programy pro vstupní i výstupní procesy, hlavní souborný katalog je provozován v systému firmy VTLS a v současnosti se pracuje na rozvoji a nasazování dalších moderních technologií pro vyhledávání a zpřístupňování informací. Mnozí odborníci z účastnických univerzit se prací na jeho dalším rozvoji aktivně zúčastňují. Celá komunita pořádá v ročních intervalech významné konference, na nichž jsou přijímána významná rozhodnutí a doporučení k perspektivním řešením toho či onoho problému [HAGEN, 2003]. URL základního sídla NDLTD (<http://www.ndltd.org/>) je velmi frekventovaným hyperlinkem na stránkách téměř všech prezentací univerzit, ale i dalších institucí a informačních systémů.

Myšlenka vzniku systému NDLTD se zrodila v roce 1987 na pracovním setkání odborníků, pořádaném známou americkou společností UMI (University Microfilms International, Ann Arbor, Michigan, <http://www.proquest.com/brand/umi.shtml>). Účastníky schůzky byli také zástupci Virginské techniky. Je zajímavostí, že záhy poté byla Y. Rubinským připravena **první verze SGML DTD** (Standard Generalized Markup Language Document Type Definition) pro strukturování e-disertace (později po zahájení provozu systému v roce 1996 byla DTD jenom mírně revidována). V letech 1989-1990 probíhaly první experimenty s konverzí dobrovolně odevzdaných disertací do připravené struktury. V roce 1992 proběhl významný seminář, na kterém 11 univerzit USA demonstrovalo zájem podílet se na elektronickém zpřístupňování disertačních a diplomových prací [FOX, 1996, část 4, Since 1987].

Byla zmíněna společnost UMI a byl zmíněn problém obtížného naplňování myšlenky systému NDLTD v samotných USA. V jakém vztahu vlastně jsou tyto dva subjekty? Dá se říci, že z určitého hlediska jde o konkurenční systémy. Je potřebné v tomto okamžiku připomenout, že UMI se, kromě jiného, po dlouhá desetiletí (od r. 1938) věnuje **registraci, archivování a komerčnímu zpřístupňování (publikování na základě objednávky)** disertačních a částečně také diplomových prací, které pocházejí prakticky ze všech severoamerických vysokých škol. V případě disertací jde o téměř 100% pokrytí. V databázi má dnes již více než 2 500 000 záznamů dokumentů (s retrospektivou od roku 1861), většina z nich je k dispozici v úplném textu ve formě mikrofilmu. Ročně přibývá do báze cca 55 000 nových záznamů, z toho cca 40 000 z USA. Základním výstupem je produkt „Dissertation Abstracts“ (zahrnuje záznamy všech předložených prací). Jeden z výstupů UMI „American Doctoral Dissertations“ zahrnuje navíc i záznamy disertací do UMI nepředložených. V tomto případě jde o národní registraci tohoto typu dokumentu v USA. Od roku 1997 se UMI věnuje plně také digitalizaci. Disertace, které byly dodány v papírové formě, jsou převáděny do podoby digitálního obrazu (ve formátu TIFF). Absolventi doktorských nebo magisterských studií, kteří předkládají své práce do UMI již v digitální formě (ve formátech PDF, PostScript, MS Word, WP), získávají pro příslušné školy jisté výhody v přístupu k produktům UMI. Předkládání disertačních prací do UMI je v USA zavedenou tradicí. Je doprovázeno příjemnou službou pro autory, totiž vyřízením autorských ochranných práv v Úřadu pro copyright při Kongresové knihovně.

V roce 1999 podepsalo UMI s americkým Úřadem pro copyright dohodu, v rámci níž se UMI stalo oficiálním repozitářem pro ukládání digitálních disertačních a magisterských prací.

Záznamy i plné texty zpřístupňuje UMI také v systému „ProQuest Digital Dissertations“. Po této krátké charakteristice činnosti UMI je zřejmé, že dosáhnout toho, aby univerzity v USA v co největším počtu změnily tuto zavedenou tradici, začaly zveřejňovat disertační práce na svých vlastních serverech zdarma a pro celý svět a připojily se tak k federaci NDLTD, není jednoduché. A to i přes to, že existují zjevné přednosti takového systému, které jsou formulovány v jeho cílech a funkcích.

Vlastní vývoj systému NDLTD začal v září 1996 po získání grantových prostředků Ministerstva výchovy USA na rozvoj 3letého projektu národní digitální knihovny disertací a diplomových prací. Připomeňme často citované nejdůležitější cíle systému. Měl zajistit **automatické předkládání disertací** (angl. submission) v elektronické formě včetně metadat ze strany autorů za podmínky dodržování stanovených standardů a metodik a měl jinak plnit běžné funkce digitální knihovny, tj. **ukládání, archivování a zpřístupňování disertací online** přes webové rozhraní. Sledovalo se dosažení **zlepšení výuky a komunikace informací produkovaných na univerzitách**. Studenti měli možnost integrovat do svých prací prvky, které tradiční forma nedovolovala (**multimédia, hypermédia** aj.). Měli si osvojit dovednosti **elektronického publikování** a tím se lépe připravit na své působení v digitální době. Očekávalo se výraznější **snížení nákladů** na celý proces předkládání a zpracování prací v univerzitních knihovnách. Výzvou ke zpřístupňování prací zdarma se předpokládal i vznešený cíl - **pomáhat studentům a odborníkům ze zemí třetího světa**. Zároveň měly být rozvíjeny **nové technologické prvky budování digitálních knihoven** [SULEMAN, 2001a].

V rámci samotné Virginské techniky, která byla jednou z prvních škol, jež zavedla povinnost odevzdávání disertací pouze v elektronické formě, byl připraven a realizován **nový pracovní postup** (angl. workflow), který je nabízen i dalším univerzitám, zejména v USA (postupy bývají podobné na rozdíl od některých jiných zemí). Zahrnuje následující podstatné momenty ze „životního cyklu e-disertace“ (<http://etd.vt.edu/>):

- ❑ **Příprava studenta** k tvorbě e-disertace a obhajobě ze strany školy (školení, styly, formáty, konverze dat aj., zajištění příruček a technické pomoci, procesy automatického předkládání práce, získání loginu do systému aj.)
- ❑ **Napsání disertace v elektronické formě** podle stanovených požadavků a předpisů; lze psát ve formátu MS WORD/WordPerfect aj.; elektronická disertace může zahrnovat i jiné typy souborů v předepsaných formátech (obrazové, video, zvukové apod.); výsledná podoba e-disertace je v časovém předstihu předávána členům komise pro obhajoby, a to buď v elektronické formě, nebo ve formě papírové, pokud si ji vyžádají
- ❑ **Obhajoba disertace** (členové komise podepisují studentům Závěrečnou zkušební kartu); dodejme, že v USA se napřed obhajuje a teprve poté předkládá konečné znění disertace
- ❑ **Závěrečné dopracování disertační práce** v elektronické formě na základě připomínek členů komise (členové komise podepisují Formulář se schválením e-disertace); student konvertuje e-disertaci do **formátu PDF**, pokud pro napsání použil běžný textový editor (povinný krok)
- ❑ **Předkládání e-disertace** (angl. submission) pomocí automatizovaného systému; proces zahrnuje vyplňování požadovaných metadat pomocí online formuláře přes WWW rozhraní a transfer všech souborů e-disertace na server školy; student zpravidla zároveň absolvuje proces předkládání disertace do UMI (viz informace výše v textu); v rámci automatického předkládání student jako autor rozhoduje o **způsobu zveřejnění nebo**

nezveřejnění e-disertace: práci lze 1. zpřístupnit okamžitě celému světu, 2. zpřístupnit pouze pro komunitu dané školy (v tomto případě se ale práce neobjeví ani v produktech UMI a ani ji nebude možné půjčovat v rámci MVS), 3. pozastavit zpřístupnění na 1 až 2 roky, například z důvodu konání patentového řízení, a pak zvolit první nebo druhou možnost a 4. zpřístupnit úplný text pro komunitu školy a zároveň poskytnout přístup k vybraným částem práce celosvětové komunitě; záznamy dvou volně zpřístupněných e-disertací (jedna je multimediální) registrovaných v systému NDLTD jsou uvedeny v seznamu použitých zdrojů [EDMINSTER, 2002a, FRANCE, 2001].

- ❑ **Kontrola souborů e-disertace** uložených na serveru školy (provádí odpovědný pracovník výpočetního centra nebo knihovny)
- ❑ **Konečné schválení e-disertace** ze strany administrativy školy (student je vyrozuměn e-mailem a získává danou akademickou hodnost)
- ❑ **Transfer metadat o e-disertaci** do katalogu knihovny školy včetně URL jejího uložení
- ❑ **Sklizení metadat o e-disertaci** pro vyhledávací procesy systému NDLTD

Po prvním roce řešení projektu systému NDLTD, od roku 1997, kdy bylo dosaženo řady dílčích úspěchů (získání členství dalších univerzit z USA i ze světa, adaptace prvního systému pro distribuované vyhledávání Dienst, jenž byl připraven pro známou digitální knihovnu NCSTRL, příprava standardů pro formáty disertací a pro metadata museli jeho představitelé zároveň řešit řadu **námitek** a **znepokojení** ze strany studentů, univerzit a vydavatelů, týkajících se především citlivých otázek autorských práv, obtížných podmínek při vytváření, konvertování a předkládání disertací podle předepsaných pokynů, stresu studentů vyvolaného striktními časovými lhůtami aj. [FOX, 1997, části 2-3].

Nicméně rozvoj systému pokračoval, a to i po skončení 3letého projektu. V říjnu 2003 měl systém celkem 190 institucionálních členů, z toho 135 univerzit (zhruba 50 % z nich bylo z USA), 7 konsorcií a 24 dalších institucí. V současnosti je již členů méně (po zavedení členských poplatků): 88 členů institucionálních z celého světa, 2 konsorcia a 2 individuální členové. Aktuální fond NDLTD čítá **767 351 metadatových záznamů dokumentů**.

Nejvíce úsilí věnuje mezinárodní komunita odborníků v současnosti rozvoji technologií vyhledávacích procesů a informačních služeb v rámci celého systému NDLTD i jednotlivých lokálních archivů. V tomto směru se dá i říci, že systém NDLTD je významným výzkumným a experimentálním pracovištěm, jehož výsledky se okamžitě uplatňují nejen v lokálních systémech jeho vlastních členů, ale i v systémech jiných typů digitálních knihoven v globálním rámci.

Historicky prvním systémem pro vyhledávání informací z NDLTD byl **systém federativního typu** (Federated Search System), v rámci něhož byl dotaz, zapsaný přes WWW formulář, odeslán směrem k indexům distribuovaných archivů [POWELL, 1998]. V další fázi vyhledávací systém zajišťoval sběr stránek s výslednými záznamy e-disertací pro potřeby jejich prohlížení ze strany uživatelů. Pomocí specifického protokolu **Dienst** bylo možné z příslušného repozitáře stáhnout úplný text disertace. Tento vyhledávací systém trpěl řadou nedostatků (ne všechny servery byly v daný okamžik k dispozici, výsledné záznamy nebyly řazeny podle relevance apod.).

Po vzniku iniciativy **OAI** v roce 1999 a později první verze **Protokolu pro sběr metadat** (OAI-PMH) v roce 2001 připravil systém NDLTD řešení k využití tohoto protokolu pro **sběr metadat** z distribuovaných repozitářů (podle definic OAI-PMH jde o poskytovatele dat) ve formátu ETD-MS [ETD-MS, 2008] a jejich konečné uložení v kódu UNICODE v databázi souborného katalogu (podle definic OAI-PMH jde o poskytovatele služeb). Sklizená metadata

byla jsou pro potřeby souborného katalogu, který začal být budován v nabídnutém novém integrovaném knihovnickém systému **VIRTUA** (The Virtua System) americké firmy VTLS (a na jejím hardwaru), shromažďována ale napřed ve speciální mezičlánkové databázi (má vlastně také charakter poskytovatele služby). Data jsou z této mezičlánkové báze automaticky importována do báze systému VIRTUA a před uložením konvertována do formátu MARC21. Systém NDLTD má v současné době ale již další nový moderní systém pro vyhledávání zvaný „VTLS Visualizer“ (<http://rogers.vtls.com:6080/visualizer/>). Zkrácený záznam disertace (a celková statistika zastoupených jazyků) je zobrazen na obr. č. 4/43. Jiným systémem pro vyhledávání je systém SCIRUS ETD Search (<http://www.ndltd.org/serviceproviders/scirus-etd-search>). Metadata jsou navíc sklížena službou „NDLTD Union Katalog“, kterou zajišťuje OCLC jednak pomocí protokolu OAI-PMH, jednak protokolu SRU.

The screenshot shows the NDLTD (Networked Digital Library of Theses and Dissertations) interface. At the top, there is a search bar with the text 'Search for items:' and a 'Search' button. Below the search bar, it says 'New search' and 'Refine this search'. The main header reads 'NETWORKED DIGITAL LIBRARY OF THESES AND DISSERTATIONS' and 'Powered by VTLS'. A blue banner indicates 'We found 767351 matching items.' and 'Collection last updated on 28/03/2009'. Below this, there is a 'Showing all items.' section and a 'Sort by: Publication Date (newest first)' dropdown menu. The main content area displays a list of results, with the first result being 'Volatiles in Melt Inclusions from Mexican and Nicaraguan Volcanoes: Implications for Complex Degassing Processes' by Atlas, Zachary D. The result includes details about the date/publisher, language, subjects, and format. A 'Show More/Less' link is also present. On the left side, there is a 'Language' section with a table showing the count of items for various languages.

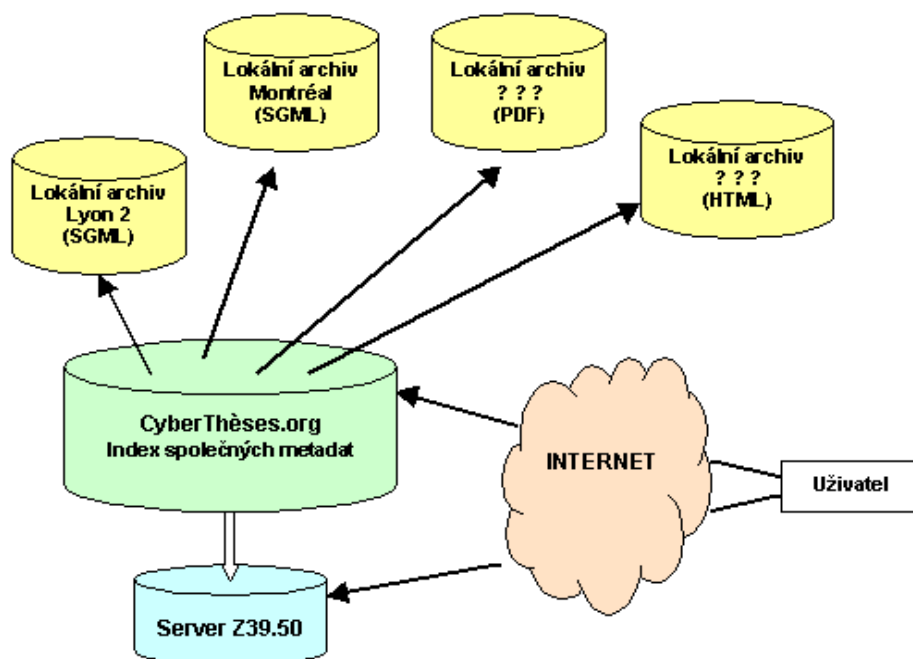
Language	Count
English	445768
Portuguese	80754
German	43948
French	25798
Chinese	20336
Lithuanian	7794
Spanish	4314
Undetermined	3988
Swedish	2942
Turkish	417
Dutch	340
Hindi	270
Japanese	266
Latin	249
Afrikaans	153
Georgian	122
Italian	105
No linguistic	105

Obr. č. 4/43: zkrácený záznam disertace fondu NDLTD v systému „VTLS Visualizer“ včetně statistiky zastoupených jazyků (vlevo) [převzato z OpenDOAR, 2009-03-25]

4.7.2.2 Mezinárodní systém e-disertací „CyberThèses“

Mezinárodní virtuální systém digitální knihovna elektronických disertací, pojmenovaný futuristicky „CyberThèses“ (Kybernetické disertace), byl založen původně v rámci frankofonního kooperačního projektu kanadské Montrealské univerzity a francouzské Univerzity Lumière, Lyon 2. Systém byl dostupný na URL: <http://www.cybertheses.org>) (dnes už ale není). Původně šlo o elektronické disertace ve francouzském jazyce. Významným rysem knihovny byla původní koncepce archivace disertací ve formátu typu SGML (konkrétně ve formátu TEI Lite) produkovaných zakládajícími univerzitami a jejich následné

zpřístupňování v různorodých formátech (SGML, HTML, XML, PDF, PostScript aj.). Zobrazování dokumentů ve formátu HTML bylo realizováno ve statickém režimu (stránky byly připraveny předem na základě konverze z formátu SGML).



Obr. č. 4/44: Schéma vyhledávání informací z aktuální verze digitální knihovny CyberThèses.org [převzato a upraveno podle dřívější dokumentace systému na WWW, 2003-10-31]

Archivační formát byl výsledkem konverze dokumentů z výchozího formátu RTF pomocí specifického programu. Elektronické disertace vytvářeli autoři podle předepsaných šablon v programech MS Word a Wordperfect. Dalším rysem systému bylo budování **společného indexu metadat** pro potřeby vyhledávání. V tomto směru se systém také otevřel dalším univerzitám světa (členy se staly další univerzity z Francie, Kanady, ale i z jiných zemí Evropy, Jižní Ameriky a Afriky, jejichž archivy mohly obsahovat elektronické disertace uložené a rozšiřované i v jiných jazycích a formátech než u dvou výchozích univerzit. Situace tehdejšího vyhledávání informací ze systému Cyberthèses je znázorněna na obr. č. 4/44. Index byl budován na základě závazného metadatového formátu, který vycházel ze standardu Dublin Core. Členové systému/sítě CyberThèses mohli pro budování lokálních archivů využívat zdarma a v plné míře řady programů, které byly v rámci systému připraveny.

Původní systém „CyberThèses“ se ale po roce 2003 proměnil. **Inovovaný systém** byl přeorientován na otevřené a volně dostupné programové zdroje a došlo k přechodu na archivační formát **XML** (celá retrospektiva elektronických disertací byla překonvertována z formátu SGML do formátu XML). Bylo využito nových nástrojů ke konverzi elektronických disertací z formátu Word do formátu XML. Prohlížení HTML stránek je realizováno na základě formátovacího jazyka **XSLT**. Byl připraven a instalován nový moderní vyhledávací systém. Systém je budován na základě budování lokálních archivů elektronických disertací a centrálních metadatových databází s jednotným vyhledáváním

Celý dnešní systém „CyberThèses“ má k dispozici dva portály:

- Portál jihoamerický na Chilské univerzitě (Universidad de Chile, Santiago), (<http://www.cybertesis.net/>) – zahrnuje přibližně 30 univerzit z celého světa (sběr metadat)

- Frankofonní portál na Lyonské univerzitě II (Université Lyon II, <http://cybertheses.francophonie.org/>)

4.7.2.3 Australasijský systém pro digitální disertace (ADT)

Jednou ze zemí, které od počátku pozorně sledovaly vznik a rozvoj systému NDLTD, je Austrálie. Nešlo však jen o sledování a získávání zkušeností od amerických partnerů. Některé významné australské univerzity brzy americké univerzity následovaly a staly se nejen přímými aktivními účastníky systému NDLTD, ale jejich odborníci začali sami realizovat myšlenku vytvoření vlastního národního systému pro rozšiřování a zpřístupňování digitálních disertací („thesí“; v Austrálii je pro doktorské disertace užíváno anglického termínu „thesis“, nikoliv „dissertation“). Jedním z faktorů iniciativy australských univerzit je i skutečnost, že v zemi neexistuje žádná celostátní registrace obhájených disertací (povinný výtisk disertací neexistuje). Národní knihovna Austrálie tyto dokumenty v rámci národní bibliografie také neregistruje. Národní souborný katalog Austrálie záznamy disertací jenom z některých fondů univerzitních knihoven, a to ne vždy v úplnosti.

Sedm největších australských univerzit, na kterých se každým rokem obhájí zhruba 2 000 disertačních prací (tj. asi 50 % z celkového počtu všech obhájených australských disertací), získalo již v roce 1997 od Rady Australského výzkumu (Australian Research Council, ARC) grantové prostředky, které jim umožnily zahájit projekt s původním názvem **Australské digitální disertace ADT** (Australian Digital Theses). Základním cílem projektu bylo vytvoření **distribuované databáze digitálních verzí disertací** produkovaných postgraduálními studenty australských univerzit a jejich zpřístupňování prostřednictvím WWW. Prospěšnou myšlenkou v pozadí celého projektu bylo zajistit přístup k australským disertacím, a tudíž i k výsledkům australského výzkumu, pro co nejširší mezinárodní komunitu (země třetího světa nevyjímaje). „Digitální verze disertací“ jsou v cíli zdůrazňovány jednak proto, že specifíkem australského programu je, oproti systému NDLTD, postupné zahrnutí retrospektivně zpracovaných dříve obhájených disertací v papírové formě, jednak proto, že disertace v papírové formě se, vedle verze digitální, musely dle požadavků a směrnic jednotlivých univerzit k obhajobě předkládat.

Projekt ADT byl realizován v letech 1998-1999, provoz systému byl pak oficiálně zahájen v roce 2000. Koordinátorem projektu byla **Rada knihovníků australských univerzit** (Council of Australian University Librarians, CAUL, <http://www.caul.edu.au>), vlastní koncepci a model systému připravili odborníci 7 partnerských univerzit, respektive jejich knihoven, které se také staly zakládajícími členy. Šlo o následující subjekty: Knihovnu Univerzity Nového Jižního Walesu (Sydney, <http://info.library.unsw.edu.au/>), Univerzitní knihovnu Melbournské univerzity (<http://www.lib.unimelb.edu.au/>), Kybernetickou knihovnu Queenslandské univerzity (Brisbane, <http://www.library.uq.edu.au/index.html>), Knihovnu Univerzity v Sydney (<http://www.library.usyd.edu.au/Home.html>), Knihovnu Australské národní univerzity (Canberra, <http://anulib.anu.edu.au/>), Knihovnu Curtinovy technické univerzity (Perth, <http://lisweb.curtin.edu.au/>) a Knihovnu Griffithovy univerzity (Brisbane-Gold Coast, <http://www.gu.edu.au/ins/lils>).

Specifíkem australského projektu je skutečnost, že celou aktivitu národního charakteru zastřešuje a koordinuje národní odborná knihovnická organizace CAUL. Lídrem projektu ADT je Knihovna Univerzity Nového Jižního Walesu (UNSW). Hlavní stránka již realizovaného systému včetně rozhraní pro vyhledávání z báze dat je lokalizována na serveru CAUL (<http://adt.caul.edu.au/>). Od roku 2000 se k projektu ADT postupně začaly připojovat další australské univerzity, takže jejich celkový počet k červnu 2002 dosáhl čísla 24. Významná změna nastala ale v roce 2004, kdy byl systém přejmenován na „**Australasian**

Digital Theses Program“ (se stejnou zkratkou ADT), a to v souvislosti s přijetím dalších členů včetně účastníků z Nového Zélandu. Systém tvoří celkem 41 členů. Od roku 2006 přešel systém na sběr metadat pro centrální databázi na protokol OAI-PMH.

Počet zaregistrovaných a uložených elektronických disertací dosáhl v dubnu 2008 hodnoty **23 634** (celkový počet záznamů v národním registru ADT byl **149 117**).

Australský program ADT je dnes ale již ve svém jádře zavedeným systémem. Lze ho charakterizovat následujícími znaky:

- ❑ Jde o otevřený informační systém národního charakteru, do kterého se již zapojila převážná část univerzit. Legislativní zajištění ze strany státu v tomto směru neexistuje. Systém vznikl z **iniciativy univerzitních knihoven** a je koordinován a rozvíjen pod patronátem **organizace australských univerzitních knihovníků**. Jde o významný rys tohoto systému. Řízení systému zajišťuje Výkonný výbor ADT. Australská národní knihovna je v systému zastoupena jako pozorovatel a poradce pro řadu specifických úloh. Lídrem projektu byla a zůstává Knihovna Univerzity Nového Jižního Walesu, která kromě jiného na svých serverových kapacitách provozuje důležité centrální agendy systému. Speciální skupiny složené z významných odborníků zajišťují další rozpracování specifických problémů (týkajících se například softwaru aj.).
- ❑ Registrovaným dokumentem je v australském systému zásadně **disertační práce** (angl. thesis) vedoucí k získání hodnosti PhD nebo hodnosti ekvivalentní. Výjimečně jsou přijímány magisterské diplomní práce, pokud byly připraveny v rámci nějakého výzkumu. Jde o účelové a v tuto chvíli rozumné a opodstatněné rozhodnutí představitelů systému. Každý doktorand, který odevzdává disertaci v digitální formě, rozhoduje jako autor vlastníci autorská práva o tom, jestli si přeje dokument zveřejnit volně přes WWW bez omezení nebo s omezením pouze pro komunitu své mateřské univerzity v intranetovém provozu. Možností je i zákaz zveřejňování přes WWW, ale nikoliv navždy. Doktorand může stanovit lhůtu nezveřejnění práce v měsících, maximálně však na 24 měsíců. Vzhledem k tomu, že v rámci systému prozatím funguje pouze dodatečné deponování již dříve obhájené, předložené a schválené disertace (v papírové formě), musí autor potvrdit autenticitu digitální kopie s kopií papírovou, která již uvedenými procesy prošla.
- ❑ V rámci systému byly připraveny a do praxe zavedeny závazné standardy. Formátem odevzdávaných digitálních a i v retrospektivě digitalizovaných disertací je zatím **formát PDF**, který má podle australských odborníků řadu výhod (snadný pro konvertování, programy pro čtení souborů jsou zdarma k dispozici, zajišťuje vysokou úroveň bezpečnosti dokumentu, je vhodný i pro účely archivace). Na univerzitách je pro doktorandy zajištěna podpora a technická pomoc. Jsou přesně zadány pokyny systému ke strukturační digitálního dokumentu i pro pojmenování dílčích souborů (využívány jsou písmenné i číselné znaky) deponované disertace. Jako povinný se online zasílá soubor s pojmenováním **„01front.pdf“** (zahrnuje názvové a autorské údaje, abstrakt, poděkování, obsah, úvod, předmluvu a případně jakýkoliv úvodní text, který není součástí těla dokumentu). Hlavní text práce má pojmenování **„02whole.pdf“** (jde-li o celý dokument). Pokud je elektronický text členěn do více částí, předávají se soubory se jmény **„02chapter1.pdf“**, **„03chapter2.pdf“**, **„04appendix.pdf“**, **„05bibliography.pdf“** apod. Pro potřeby doktorandů je pro vytváření metadat (standard ADT) k dispozici pohodlný interaktivní formulář k zápisu a generování metadat. Byly stanoveny přesné pokyny pro zápis a kódování znaků, zejména pro případ zápisu symbolů a diakritiky.
- ❑ Systém ADT zahrnuje, což je opět typické, retrospektivní zpracování disertací, které byly obhajovány v minulosti. V tomto případě je zajištěna nutná adaptace základního softwaru převzatého ze systému NDLTD. V první řadě se digitalizují disertace, o které je velký

zájem v rámci výpůjčních služeb. Byl připraven kvalitní postup digitalizačního procesu a odpovědnou organizací v tomto směru je Knihovna Melbournské univerzity. Zabezpečuje poradní služby a technickou pomoc pro další účastnické organizace programu ADT.

- Architektura systému ADT je založena na distribuovaném ukládání digitálních disertací i metadat na serverech jednotlivých účastnických univerzit a na centrální tvorbě báze metadat (pomocí protokolu OAI-PMH) a centrálním vyhledávacím systémem zajišťovaném kapacitami serveru organizace CAUL.

4.7.2.4 Národní německý systém „DissOnline.de“

Také německé univerzity patří k institucím, které ve druhé polovině 90. let 20. století zahájily zkoumání možností rozvoje elektronického publikování svých vlastních produktů, mezi které patří zejména vysokoškolské kvalifikační práce (něm. Hochschulschriften). Řada známých univerzit v rámci svých WWW prezentací postupně začala nabízet zpřístupňování úplných textů elektronických disertací na WWW (s otevřeným nebo uzavřeným přístupem pro veřejnost). Některé z univerzit se zapojily do systému NDLTD a jejich odborníci se podílejí neustále intenzivně na řešení mnoha aktuálních úkolů.

Pokud jde o **bibliografickou registraci disertací**, pak v SRN ji plně a spolehlivě zajišťuje **systém národní bibliografie** (Deutsche Nationalbibliographie). Bibliografická registrace je v péči Německé národní knihovny (Die Deutsche Nationalbibliothek), která na základě nařízení o odevzdávání povinného výtisku z roku 1982 (Verordnung über die Pflichtablieferung von Druckwerken an die Deutsche Bibliothek : Vom 14. Dezember 1982) disertační ale i habilitační práce v tradiční, ale dnes i elektronické formě (od 1.2. 1998) získává a registruje. Záznamy jsou pak dostupné prostřednictvím nejrozličnějších výstupů: papírových či mikrofišových s tradičním názvem „Reihe H, Hochschulschriftenverzeichnis“ nebo na CD-ROM, který je publikován pod názvem „Diss-CD“ a zahrnuje retrospektivu od roku 1945 s celkovým objemem téměř 1 000 000 záznamů. Každý doktorand v SRN **je povinen doktorskou práci nějakým způsobem publikovat**. Pro mladé doktorandy znamenal tradiční tisk disertace vynaložení značného objemu finančních prostředků. I to byl jeden z důvodů, který inicioval založení národního projektu na zpracování a zpřístupňování disertací v digitální formě.

První významný projekt vznikl v SRN nejprve na Humboldtově univerzitě v Berlíně (Humboldt Universität zu Berlin) na základě společného úsilí jejího výpočetního centra a knihovny [DIEPOLD, 1999]. Bezprostředním iniciátorem projektu byla pracovní skupina „IuK Initiative“ Německých vědeckých společností (Initiative Information und Kommunikation der wissenschaftlichen Fachgesellschaften in Deutschland). Projekt byl sponzorován známou Německou výzkumnou nadací (Deutsche Forschungsgemeinschaft, DFG, <http://www.dfg.de>).

První etapa projektu, který získal název „**Dissertationen Online**“, probíhala od jara 1998 do března 1999. Druhá etapa, grantovaná také DFG v celkovém objemu 300 000 EU, probíhala od března 1999 do března 2000 a účastnila se jí již větší skupina spolupracujících výpočetních center univerzit a jejich knihoven a také Německá (národní) knihovna. Lídrem projektu se stala Humboldtova univerzita v Berlíně a hlavním řešitelem projektu se stal prof. Peter Diepold z Institutu pro hospodářskou pedagogiku a vzdělávání dospělých téže univerzity. Základním cílem bylo vybudování národní **virtuální digitální knihovny**, v rámci níž by německé disertační a jiné kvalifikační práce byly uloženy, archivovány a zpřístupňovány v národním i mezinárodním měřítku. Úsilí odborníků, kteří se na projektu podíleli, bylo zaměřeno na:

- vysokou kvalitu podpory autorům-doktorandům (pomocí speciálních kursů a příruček pro tvorbu e-disertací a speciálních služeb pro přípravu grafických a zvukových informací)
- navržení modelu pro tok informací v rámci univerzity a univerzitní knihovny (od předkládání e-disertací, přes jejich zpracování a uložení v digitálním repozitáři až k jejich zpřístupňování přes WWW); k tomu účelu byl pro automatizovanou katalogizaci připraven původní formát MetaDiss založený na standardu Dublin Core, který později nahradil formát XMetaDiss; pro autory se předpokládalo vytvoření pohodlného online formuláře pro tvorbu metadat; v systému se předpokládala konverze všech disertací do formátu XML pro potřeby jejich trvalého uložení a archivace
- dlouhodobou archivaci a ochranu elektronických disertací ve formátu XML; předpokládala se příprava typické struktury dokumentu pomocí DTD (informace o DiML.DTD); v rámci archivace se počítalo s využitím kódu kontrolního součtu digitálního objektu
- implementaci vysoce kvalitního vyhledávacího systému, který by integroval vyhledávání v metadatech, úplném textu i struktuře dokumentu; systém by zahrnoval i inteligentní prohlížení celého fondu za využití tezauru a více druhů třídících schémat.

Již během prvního roku se podařilo získat rozhodnutí Akademického senátu Humboldtovy univerzity, na jehož základě bylo možné dobrovolné předkládání disertací v elektronické formě jako jedné z publikačních možností. Pro doktorandy byla připravena školení. Byl vyvinut a zpřístupněn speciální program pro vyplňování metadat – **interaktivní formulář MMMfT** (My Meta Maker for Theses). Uložené disertace získaly kódy kontrolního součtu digitálního objektu a byly archivovány na certifikovaném dokumentovém serveru. Byla navržena první DTD pro elektronické disertace (DiML, později vznikl formát XDiML) a také odpovídající šablony pro editor MS Word (ve spolupráci s odborníky z ND LTD).

Dvouletý projekt „Dissertationen Online“ byl ukončen úspěšnou oponenturou ve dnech 30.-31. října 2000 v Berlíně. Oponentura neznamenal definitivní ukončení projektu, stala se odrazovým můstkem k dalšímu pokračování, tentokrát již pod patronátem Německé knihovny (od února 2001) ve společném projektu nazvaném „**DissOnline.de**“ [DIEPOLD, 2001, Entwicklung]. Název projektu vyjadřuje zároveň URL jeho základního webového sídla (<http://www.dissonline.de>).

Studenti německých univerzit mohou v současné době předkládat disertaci v několika publikačních formách (<http://www.dissonline.de/>): 1. na mikrofiši, 2. ve formě seznamu publikovaných vědeckých článků v odborných časopisech, 3. ve formě 40-150 papírových exemplářů, 4. ve formě publikace vydané vydavatelstvím (často univerzitním) s příslibem alespoň průměrného nákladu a 5. v elektronické formě, která ovšem musí být doplněna 4-8 papírovými exempláři. Disertaci v elektronické formě ale nepřijímají v současné chvíli všechny německé univerzity. Záleží na aktuálním statutu jejich řádů doktorských studií.

Většina německých univerzitních knihoven využívá v rámci národního projektu DissOnline pro plnotextové vyhledávání informací ze svých lokálních archivů různé programové aplikace (softwaru). Národní systém má již také centrální kvalitní a sofistikovaný vyhledávací systém (<http://search.dissonline.de/>), který vyhledává záznamy nad centrální metadatovou databází (data jsou sklízena přes protokol OAI-PMH). Původní metadatová specifikace METADISS byla později nahrazena specifikací novou XMetaDiss (je založena již plně na struktuře XML), která se vyznačuje vysokou profesionalitou. Záznam disertace získaný z repozitáře Humboldtovy univerzity v Berlíně přes protokol OAI-PMH (<http://edoc.hu-berlin.de/OAI-2.0>) v tomto formátu (oai_xmetadiss) je k dispozici na obrázku v další 5. kapitole textu.

Německé disertace jsou opatřovány jedinečným trvalým identifikátorem URN v rámci celonárodního systému, který spravuje Německá národní knihovna ve Frankfurtu nad Mohanem. Každá disertační práce v elektronické podobě je ukládána jednak v repozitáři mateřské univerzity, jednak v centrálním depozitním skladu Německé národní knihovny v systému společnosti IMB „DIAS“ (zabezpečení dokumentu je dokonalé, existuje až 6 digitálních kopií každého dokumentu lokalizovaných na různých serverech v Německu).

4.7.2.5 Situace s rozšiřováním vysokoškolských kvalifikačních prací v České republice

Zhruba po roce 2001 se začaly první české univerzity a jejich univerzitní knihovny zajímat o zpracování a rozšiřování disertačních a jiných vysokoškolských kvalifikačních prací [ŘEPIŠOVÁ, 2000]. První lokální systémy vznikly na Ostravské univerzitě, Zlínské univerzitě, Masarykově univerzitě aj. [BRATKOVÁ, 2004b]. Tyto systémy se vyznačovaly implementací různorodých softwarů, ale zejména různorodostí metodických postupů při zpracování VŠKP. Chyběl jednotný metadatový standard, chyběly další nástroje. Diskutovali se autorsko-právní problémy zpřístupňování prací. V období 2002/2003 se proto na platformě knihovnické organizace SKIP vytvořil Klub vysokoškolských knihovníků, který začal problematiku zpřístupňování elektronických VŠKP diskutovat v širším plénu. Později v roce 2004 byla v rámci Asociace knihoven vysokých škol založena, jak bylo v úvodu této podkapitoly uvedeno, Odborná komise pro otázky elektronického zpřístupňování VŠKP (<http://www.evskp.cz/>), která začala uvedené otázky řešit. Velkým požadavkem řady vysokých škol bylo zejména zpracování jednotného metadatového standardu – též pro potřeby budování národního metadatového registru. Dílčí pracovní skupina tento standard připravila (od roku 2004) a v roce 2005 zveřejnila jeho verzi 0.1 (vývojová verze [EVSKP-MS, 2005]). Jeho další vývoj měl již být zajištěn grantovou činností, což se podařilo až v roce 2008, nejdříve ve verzi 1.0 [EVSKP-MS, 2008-05-05] a poté ve verzi 1.1 [EVSKP-MS, 2008-07-15]. Tato otázka je předmětem další části disertační práce k kapitole 5.

Odborná komise připravila a na svém webu zveřejnila i řadu dalších doporučení. Do procesu zpracování VŠKP v elektronické formě se zapojily i další české vysoké školy. Začala být diskutována otázka vytvoření (po vzoru ze zahraničí) národního metadatového registru VŠKP. Tato myšlenka se stala reálnou s přijetím Centralizovaného projektu „Národní registr VŠKP a úložiště závěrečných prací se službou na odhalování plagiátů“, který podpořilo MŠMT (MŠMT C1/2008). Řešitelem projektu se stala Masarykova univerzita v Brně. Na základě dohody Odborná komise pro otázky elektronického zpřístupňování VŠKP dostala za úkol připravit výše zmíněné národní metadatové standardy jako výměnné formáty.

Centralizovaný projekt řešený v roce 2008 je v základu vyřešený, systém národního registru je založen a provozován Masarykovou univerzitou (<http://theses.cz/>). Kromě role registrační plní ještě úlohu systém pro odhalování plagiátů. Pokud jde o registr, je nutné podotknout, že se na jeho budování zatím nepodílejí všechny veřejné vysoké školy ČR (chybí ČVUT, UK v Praze), takže prozatím nelze hovořit o 100% registrační databázi pokrývající produkci všech českých vysokých škol. K účastníkům patří:

1. Akademie múzických umění v Praze
2. Brno International Business School
3. Česká zemědělská univerzita v Praze
4. Ekonomická univerzita v Bratislavě

5. Janáčkova akademie múzických umění v Brně
6. Jihočeská univerzita v Českých Budějovicích
7. Masarykova univerzita
8. Ostravská univerzita v Ostravě
9. Slezská univerzita v Opavě
10. Univerzita Hradec Králové
11. Univerzita Jana Evangelisty Purkyně v Ústí nad Labem
12. Univerzita Palackého v Olomouci
13. Univerzita obrany
14. Univerzita Tomáše Bati ve Zlíně
15. Vysoká škola báňská Technická univerzita Ostrava
16. Vysoká škola ekonomická v Praze
17. Vysoká škola finanční a správní
18. Vysoká škola polytechnická Jihlava
19. Vysoká škola technická a ekonomická v Českých Budějovicích
20. Vysoká škola uměleckoprůmyslová v Praze
21. Západočeská univerzita v Plzni

V současné době je k dispozici starší rozhraní pro vyhledávání informací o VŠKP, nový systém se připravuje. Rozšíření systému o další školy se předpokládá v dalších letech.

5 Identifikace a popis digitálních objektů v digitálních knihovnách

Tato kapitola disertační práce má praktický a také syntetický charakter. Autorka v ní po úvodním představení významných metadatových schémat užívaných v současných digitálních knihovnách představuje, v návaznosti na předchozí kapitolu 4 a zejména podkapitulu 4.7 věnovanou systémům pro zpřístupňování elektronických vysokoškolských prací v elektronické formě, výsledky vlastní výzkumné a vývojové činnosti v oblasti metadatových standardů. Představeny budou výsledky analýzy vybraných zahraničních metadatových schémat pro popis a komunikaci vysokoškolských prací a dále návrhy a reálné verze českých národních metadatových standardů pro popis VŠKP a standardů souvisejících pro popis fyzických osob a korporací. K této části práce se váží přílohy 1-4 umístěné na konci práce. Jde o vlastní výsledek (produkt) výzkumné a vývojové činnosti autorky (v grantové úloze FRVŠ č. 2045/2003 (okruh E2) „E-DI“ (Digitální knihovna diplomových a disertačních prací UK v Praze) a v grantové úloze Centralizovaný projekt „Národní registr VŠKP a úložiště závěrečných prací se službou na odhalování plagiátů“, MŠMT C1/2008). Připojené dokumenty byly vytvořeny ve spoluautorství J. Macha (Vysoká škola ekonomická). Autorčin podíl je vymezen především **oblastí sémantiky představovaných metadatových standardů**. Dlouhodobě se zabývala v profesionální a výukové oblasti bibliografickými formáty dat včetně metadat. Připojeny budou ještě části související problematiky věnované otázkám trvalých identifikátorů dokumentů i fyzických osob a korporací. Uvedená témata autorka v předchozím období zveřejnila v sérii publikovaných dokumentů.

5.1 Metadata

Metadata jsou často definována jako „data o datech“, ale autorka ne zcela s takovou definicí souhlasí, v řadě případů nepůjde o data o datech, pomocí metadat lze dnes popisovat jakékoliv informační objekty, elektronické i neelektronické, konkrétné předměty, osoby aj. a také abstrakce. Lepší definici poskytuje například směrnice připravená K. M. Wissorovou v rámci grantové úlohy k digitalizaci [Guidelines, 2007]:

Metadata jsou strukturované informace, které popisují, vysvětlují, lokalizují nebo jinak pomáhají snadněji vyhledat a využít informační zdroj. Metadata se dnes, i v kontextu budování a provozu digitálních knihoven rozděluje na:

- **Administrativní metadata**, která jsou primárně určena k usnadnění správy informačních zdrojů
- **Deskriptivní metadata**, která jsou primárně určená pro potřeby zjišťování, identifikace a vyhledávání informačních objektů
- **Technická metadata**, která jsou primárně určena k vytváření a charakteristice počítačových souborů
- **Strukturální metadata**, která popisují interní organizaci informačního zdroje a jeho umístění v nějaké externí organizaci včetně vazeb na jiné informační zdroje
- **Právní metadata**, určená primárně pro správu práv k informačnímu zdroji (jde o typ administrativních práv)
- **Archivační metadata**, určená pro správu procesů zajišťujících dlouhodobou archivaci a ochranu použitelnost digitálních informačních zdrojů

Metadatovým schématem je míněn soubor metadatových prvků (údajů) a pravidel pro jejich využití; schéma bývá definováno pro zvláštní účel.

V oblasti digitálních knihoven se v současné době využívá větší množství metadatových specifikací a metadatových schémat lokálního i mezinárodního charakteru. Trendem v oblasti vývoje digitálních knihoven jsou metadatová schémata připravená na bázi standardu XML [Gartner, 2008]. V oblasti popisných metadat patří k nejznámějším a nejvyužívanějším specifikacím **Dublin Core**, kterou spravuje mezinárodní komunita DCMI, a specifikace **MODS**, kterou spravuje Kongresová knihovna. V oblasti administrativních, právních a strukturálních metadat se uplatňuje specifikace **METS**, kterou také spravuje Kongresová knihovna.

5.2 Metadatové standardy pro VŠKP

5.2.1 Metadata pro elektronické disertace

Existující systémy pro rozšiřování elektronických disertací využívají různých souborů metadat, definovaných na lokální, národní nebo mezinárodní úrovni. Počty prvků, jejich kvalifikátorů a návazných schémat pro stanovení jejich hodnot se liší, lze však konstatovat, že většina z nich vznikla a dotváří se na základě specifikace metadatových prvků Dublin Core (DC, Dublin Core Metadata Element Set). Některé specifikace jsou co do počtu prvků menší než DC, jiné větší. Některé, jež souvisejí s národní bibliografickou registrací, bývají i rozsáhlejší a zahrnují zpravidla také prvky, které ve specifikaci DC nejsou a ani pravděpodobně nikdy nebudou. Typickým rysem všech souborů metadat pro elektronické disertace a jiné kvalifikační práce však je, že se snaží uplatnit prvky pro ně zcela typické, jako jsou například typ obhajované práce, přidělovaná akademická hodnota, datum obhajoby, instituce přidělující tuto hodnotu aj.

Následující analýza je věnována 3 vybraným specifikacím metadat, které byly připraveny pro systémy národní a mezinárodní úrovně (platné v roce 2004). Ukáže nejen na rozdílné počty prvků a jejich kvalifikátorů, ale v souvislosti s tím i na potřeby institucí a systémů, které metadata společně s elektronickými disertacemi zpracovávají, ukládají a zpřístupňují, a také na problémová místa jednotlivých definic, která mohou působit potíže při výměně dat. Jak je patrné z obr. 5/1, vybrané a porovnávány jsou specifikace tří známých metadatových souborů pro popis disertací. Jde o:

- a) co do počtu prvků nejmenší soubor **Standardizovaných metadat australského programu pro digitální disertace** (Australian Digital Theses Program Metadata Standard), který využívá převážná část univerzit v Austrálii (ve verzi z roku 2003)
- b) pro tuto analýzu základní **Soubor metadat ETD-MS** (Electronic Theses and Dissertations - Metadata Set), který využívají zejména americké univerzity sdružené v systému NDLTD (ve verzi 1.00, 2. revizi z 3. září 2003) a
- c) zcela specifický soubor metadat pro vysokoškolské práce **METADISS** (Format des Metadatenatzes für Online-Hochschulschriften), který byl společným formátem Německé knihovny a projektu SRN „Dissertationen Online“ (ve verzi 1.5 z 5. května 2003).

Z úplných definic vybraných souborů metadat jsou na obr. č. 5/1 uváděny pouze identifikátory prvků, jejich kvalifikátory (v angličtině) a v určitých případech také možná schémata pro stanovení hodnot prvků nebo kvalifikátorů. V tabulce nejsou uváděna oficiální jména prvků ani jejich definice, komentáře nebo jiné informace. Pořadí prvků je zcela dodrženo u německého a amerického souboru, u australského je pořadí několika prvků změněno oproti původnímu seznamu tak, aby bylo možné lépe prvky porovnávat. Identifikátory uváděné kurzívou náleží prvkům nepovinným. Opakovatelnost prvků z původních definic není v tabulce označována (převážná část prvků ale tuto vlastnost má).

Porovnání tří metadatových souborů se souborem DC (ten není součástí tabulky, předpokládá se ale jeho všeobecná znalost (<http://dublincore.org/documents/2003/06/02/dces/>) a také mezi sebou navzájem vede k následujícím hlavním závěrům:

1. **Australský soubor** má pragmatický ráz a je připraven především s ohledem na snadné využívání ze strany autorů, kteří samostatně pomocí editorů metadata připravují. Ve srovnání s dalšími specifikacemi má nejméně prvků i kvalifikátorů. Dodržuje specifikaci DC, i když řadu z jeho prvků (z 15 celkem 6) zcela vypouští (zařazeny nejsou prvky s identifikátory DC.contributor, DC.type, DC.format, DC.source, DC.relation a DC.coverage). Jednoduchost specifikace se projevuje v minimálním zařazování kvalifikátorů, což odpovídá současnému směru definice DC. Poněkud překvapující je nezařazení prvku „Zdroj“ (Source), protože australský program je silně orientován i na retrospektivní zpracování „papírových“ disertací. Jak uvádějí v připomínkové zprávě ke specifikaci metadat systému NDLTD (ETD-MS) odpovědní pracovníci programu ADT, nadbytečné se jim zdálo zařazovat kvalifikátor například pro prvky „Název“ (DC.title.alternative) nebo „Popis“ (DC.description.note, DC.description.release).
2. Soubor metadat nejvýznamnějšího, dnes již mezinárodního systému **NDLTD**, který je ve své dnešní verzi výsledkem kolektivní tvorby odborníků z různých zemí světa, vynechává v současné verzi ze sestavy DC pouze 2 prvky, a sice údaj o původním zdroji digitalizované práce (identifikátor DC.Source), protože NDLTD není orientován na retrospektivní zpracování, a údaj o vazbách e-disertací na související dokumenty (identifikátor DC.Relation). Vynechány jsou i některé kvalifikátory ze souboru DC, například několik kvalifikátorů pro identifikátor DC.Date nebo DC.Format. Z druhé strany tvůrci souboru ETD-MS usoudili, že je nezbytné některé prvky a určité kvalifikátory k již zařazeným prvkům doplnit. Zcela nově do sestavy ETD-MS přibyl prvek pro bližší určení získávané akademické hodnosti, který je označen vlastním identifikátorem metadat ETD-MS „thesis.degree“, i jeho 4 kvalifikátory pro určení jména hodnosti (thesis.degree.name), typu studijního programu (thesis.degree.level), studijního oboru (thesis.degree.discipline) a instituce tuto hodnost přidělující (thesis.degree.grantor). Novými kvalifikátory, jež nejsou součástí definice DC, jsou také kvalifikátory prvků „Popis“ (dc.description.note a dc.description.release) a „Příspěvatel“ (dc.contributor.role) s tím, že v hodnotě role se mohou vyskytovat různé výrazy: kromě v tabulce uvedené role „referee“ (oponent) a „advisor“ (školitel) je možné užít hodnot „committee member“ (člen komise pro obhajobu), „chair“ (vedoucí komise) aj. Na uvádění údajů zejména o školiteli/oponentech jsou v mezinárodním rámci rozdílné názory, soubor ETD-MS je nakonec zařadil, nejsou ale povinné. Problém by mohl nastat při jejich převodu do jiných systémů digitálních knihoven nebo nadstavbových servisních služeb.
3. Německá specifikace **METADISS** je velmi propracovaným souborem metadat pro popis vysokoškolských kvalifikačních prací. Ve srovnání s australskou i americkou specifikací je již na první pohled vidět, že je obohacena o řadu prvků dodaných pro potřeby německé národní bibliografie. Základem specifikace je opět metadatový soubor DC, ale zahrnuje také několik zcela nových prvků označovaných identifikátorem „DDB“, například údaj o e-mailu kontaktní osoby za instituci, která získává a zpracovává e-disertaci (DDB.Contact), nebo údaj o instituci a místu, na jejímž serveru je e-disertace uložena (DDB.Server). Takové údaje budou sloužit pro potřeby národního bibliografického systému. V německém souboru, který vychází plně ze souboru DC, je v současné verzi vynechán jediný prvek s identifikátorem DC.Coverage. Bohaté je zastoupení kvalifikátorů, dokonce i některých z dřívější specifikace DC, které jsou dnes zrušeny (například kvalifikátory DC.Creator.PersonalName, DC.Publisher.CorporateName aj.). V rámci nové verze formátu METADISS 1.5 z roku 2003 byla porvé definována již

i samostatná množina autoritních metadat (METAPERS, Version 1.1, http://deposit.ddb.de/metadiss.htm#top_II). Její specifikace odpovídá podrobným údajům o tvůrcích ze specifikace bibliografické (viz například kvalifikátor DC.PersonalName.DateOfBirth). Oproti aktuální verzi souboru DC je v německé sestavě metadat i řada vlastních kvalifikátorů, například DC.Title.Translated, DC.Contributor.Advisor, DC.Date.Accepted aj. Tyto snahy však vycházejí z potřeby mít o e-disertacích přesnější informace, jež jsou pro tento typ dokumentů specifické. Německá definice zahrnuje stále také ještě kvalifikátor LANG, který již dlouho není v souboru DC přítomen.

4. Při pohledu na údaje na obr. 1/5 se ukazuje, že 3 vybrané soubory metadat se zcela shodují (pomineme-li různá užívaná kvalifikační schémata) pouze v prvku „**Předmět**“ (DC.Subject), který je všude i povinný, v prvku „**Identifikátor**“ (DC.Identifier), který je také všude povinný, v prvku „**Jazyk**“ (DC.Language), který není povinný u americké specifikace, a v prvku „**Práva**“ (DC.Rights), který je povinný pouze v australské sestavě. V rámci prvků „**Předmět**“ se u všech specifikací uplatňují různá schémata (SCHEME), australský soubor v současné době uplatňuje volně tvořená klíčová slova, plánuje se však nasazení hesláře LCSH. Německý soubor i s ohledem na národní bibliografii zahrnuje tři pořadací jazyky: vlastní třídění německé národní bibliografie (DNB-Sachgruppen), soubor předmětových autorit (SWD) a volně i další známé třídící systémy (například Deweyho třídění, které se právě v SRN překládá, aj.). Pokud jde o **identifikátory**, pak se u všech upřednostňuje zejména trvalý identifikátor URN nebo jeho obecný rámec URI, soubor německý pak zařazuje i další typy identifikátorů. Německý soubor jako jediný uplatňuje i identifikátory pro části (díly) dokumentů. V rámci prvku „**Jazyk**“ se uplatňuje především kódování podle mezinárodní normy ISO 639-2, australský soubor a také soubor ETD-MS však také počítají s internetovou normou RFC 1766. U dalších prvků všech tří metadatových souborů nastávají menší i větší rozdíly, u některých je však i větší nesoulad. Zajímavostí je, že některé ryze specifické znaky e-disertací jsou u těchto vybraných souborů metadat zařazovány vždy do jiného prvku.
5. Mírný nesoulad se týká **názvových údajů**. Hlavní název je všude jako povinný údaj, americká i německá specifikace však využívají pro uvádění podnázvů nebo dalších názvů (jako údajů nepovinných) kvalifikátor „DC.Title.Alternative“, specifikace australská tyto názvové údaje zařazuje k názvu hlavnímu za použití příslušné interpunkce (údaje se uvádějí podle titulní stránky). Německý soubor navíc, z důvodů zřejmých, obsahuje vlastní nepovinný kvalifikátor pro překlad názvů německých prací do jiných jazyků (DC.Title.Translated a DC.Title.Alternative.Translated). Tento prvek se ale nově objevil i ve výčtu opravené verze specifikace ETD-MS (obr. 5/1). Pokud jde o **přímé tvůrce** prací (autory), počítají všechny sestavy metadat, z důvodů taktéž zjevných, s autorem jediným. U všech jde o údaj povinný, odlišnost se ale týká formy zápisu jmen autorů. Německá sestava upřednostňuje stále ještě zápis jména autora ve formalizované podobě. Uplatňuje se v kvalifikátoru DC.Creator.PersonalName (autor je zapisován v modelu: Příjmení, Křestní jméno). Americká sestava se striktně drží aktuální definice DC. Jméno autora se zapisuje podle dokumentu, tj. zpravidla v přirozeném pořadí. Totéž stanovisko se objevilo i v nově revidované sestavě australských metadat. Toto stanovisko je v současné době již podporováno také plánovaným automatickým propojováním na autoritní záznamy pomocí přiděleného URI, v němž by formalizovaný tvar mohl být, vedle jiných, uveden. Německá specifikace jde ještě dále a definuje další kvalifikátory pro bližší údaje o autorech (datum a místo narození a adresa).

Obr. č.1/5: Srovnání základních prvků australské, americké a německé specifikace metadat pro popis e-disertací a jiných kvalifikačních prací (údaje uvedené kurzívou jsou definovány jako nepovinné)

ADT Metadata	ETD-MS	METADISS
		<i>DC.Publisher.CorporateName</i>
		<i>DC.Publisher.CorporateName.Address</i>
DC.title	dc.title	DC.Title LANG="..."
	<i>dc.title.alternative</i>	<i>DC.Title.Alternative LANG=" "</i>
		<i>DC.Title.Translated LANG=" "</i>
		<i>DC.Title.Alternative.Translated LANG=" "</i>
DC.creator	dc.creator	
		DC.Creator.PersonalName
		<i>DC.Creator.PersonalName.DateOfBirth</i>
		<i>DC.Creator.PersonalName.PlaceOfBirth</i>
<i>DC.creator [email]</i>		<i>DC.Creator.PersonalName.Address</i>
DC.subject [scheme="LCSH"]	dc.subject SCHEME=" "	
		DC.Subject SCHEME="DNB-Sachgruppe"
		<i>DC.Subject SCHEME="/BK, DDC, ...]"</i>
		<i>DC.Subject SCHEME="/SWD, MeSH, ...]"</i>
DC.description (abstract only)	<i>dc.description</i>	<i>DC.Description LANG=" "</i>
		<i>DC.Description SCHEME="URL"</i>
DC.description.abstract	<i>dc.description.abstract</i>	
	<i>dc.description.note</i>	<i>DC.Description.Note LANG=" "</i>
	<i>dc.description.release</i>	
DC.publisher (school)	<i>dc.publisher</i>	
		DC.Publisher.CorporateName
		DC.Publisher.CorporateName.Address
		DDB.Contact.ID
		DDB.Contact CONTENT="[E-mail]"
	<i>dc.contributor</i>	
		DC.Contributor.CorporateName
		<i>DC.Contributor.CorporateName.Address</i>
		<i>DC.Contributor.CorporateName.Address</i>
		<i>DC.Contributor.CorporateName.Address</i>
	<i>dc.contributor.role [referee]</i>	<i>DC.Contributor.Referee</i>
	<i>dc.contributor.role [advisor]</i>	<i>DC.Contributor.Advisor</i>
DC.date scheme="W3CDTF"	dc.date SCHEME="ISO8601"	
		DC.Date.Accepted SCHEME="W3CDTF"
		<i>DC.Date.Submitted SCHEME="W3CDTF"</i>
		<i>DC.Date.Created SCHEME="W3CDTF"</i>
	dc.type	DC.Type
		DC.Identifier SCHEME="URL"
	<i>dc.format SCHEME="IMT"</i>	DC.Format SCHEME="IMT"
		<i>DDB.Server</i>
		<i>DDB.Description.Note</i>
		<i>DDB.Identifier.Transfer SCHEME="URL"</i>
DC.identifier scheme="URI"	dc.identifier SCHEME="URI, URN"	<i>DC.Identifier SCHEME="URN:NBN:DE, URN, DOI"</i>
		DDB.Identifier.Fingerprint SCHEME="MD5"
		<i>DC.Identifier SCHEME="URN" [Teil]</i>
		<i>DC.Identifier SCHEME="URL" [Teil]</i>
		<i>DDB.Identifier.Fingerprint SCHEME="MD5" [Teil]</i>
		DC.Date.Delivery SCHEME="W3CDTF"
		<i>DC.Source</i>
		<i>DC.Source SCHEME="URL"</i>
DC.language scheme="RFC3066"	<i>dc.language</i> SCHEME="ISO639-2, <i>RFC1766"</i>	DC.Language SCHEME="ISO639-2"
		<i>DC.Relation</i>
		<i>DC.Relation SCHEME="URL"</i>
	<i>dc.coverage</i>	
DC.rights	<i>dc.rights</i>	<i>DC.Rights</i>
		<i>DC.Rights SCHEME="URL"</i>
	<i>thesis.degree.name</i>	
	<i>thesis.degree.level</i>	
	<i>thesis.degree.discipline</i>	
	<i>thesis.degree.grantor</i>	

6. Větší rozdíly se vyskytují u prvku „**Příspěvatel**“ (DC.Contributor). Australský soubor tento prvek zcela vylučuje s odůvodněním, že není třeba registrovat další původce (školitele a oponenty). Poznamenejme, že údaje o členech komisí pro obhajoby ale autoři mohou (nepovinně) vyplňovat v online formuláři při deponování elektronického textu disertace (<http://www.library.unsw.edu.au/thesis/adt-ADT/info/example.html>). Americká specifikace tento prvek jako nepovinný zahrnuje, ale oproti definici ze souboru DC zavedla nový kvalifikátor dc.contributor.role s nabídkou dílčích hodnot „referee“ (oponent), „advisor“ (školitel), „chair“ (předseda) aj. Toto rozhodnutí bylo asi ovlivněno odborníky ze SRN, kteří participují v systému NDLTD. Poukazuje na to i skutečnost, že dvě uvedené role se vyskytují v německém souboru METADISS v podobě kvalifikátorů DC.Contributor.Referee a DC.Contributor.Advisor, jež jsou nepovinné. Zcela ojedinělý je ale v souboru METADISS v rámci tohoto prvku povinně uváděný kvalifikátor DC.Contributor.CorporateName, který obsahuje jméno instituce (vysoká škola, fakulta), v rámci které se práce předkládá a která také přiděluje akademickou hodnotu. Pro srovnání, australský soubor má tuto důležitou informaci obsaženou v identifikátoru DC.Publisher a americký soubor ETD-MS v rámci identifikátoru kvalifikátoru thesis.degree.grantor. Tato nesrovnalost představuje problém pro případnou výměnu metadat mezi systémy. Nepovinný prvek „**Vydavatel**“ (DC.Publisher) je v americkém souboru ETD-MS věnován údajům o instituci, která e-disertaci zpřístupňuje (popřípadě digitalizuje nebo archivuje), a doporučuje se, aby byl uváděn v přirozeném tvaru dle dokumentu. Podobně je v německém souboru METADISS v pořadí druhý identifikátor kvalifikátor DC.Publisher.CorporateName (povinný) věnován instituci, která zajišťuje zpřístupňování e-disertace (například univerzitní knihovna, výpočetní centrum apod.). Stejný identifikátor (v pořadí první) je v německém souboru vyhrazen pro údaj o národní bibliografické agentuře, která záznam metadat přebírá a dále zpracovává pro vlastní potřeby.
7. V rámci prvku „**Popis**“, který je povinný pouze u australského souboru, se uplatňuje zejména abstrakt, popř. resumé. ETD-MS definuje uplatnění textu abstraktu buď v nekvalifikovaném prvku (identifikátor dc.description), nebo v kvalifikovaném (dc.description.abstract), německý soubor v základním prvku doplněném o kvalifikátor jazyka (DC.Description LANG) - doporučuje se uplatnění vícejazyčného resumé. Australský soubor definovala (v roce 2003) buď jenom nekvalifikovaný prvek (DC.description) pro zápis abstraktu, nebo připouští stále ještě prvek kvalifikovaný (DC.description.abstract). Předmětem výtky australských odborníků je zařazení dalších kvalifikátorů v souboru ETD-MS, a sice dc.description.note (pro zápis poznámky) a dc.description.release (pro údaj o verzi práce), protože nejde o schválené kvalifikátory v souboru DC. Vlivem toho se zřejmě ocitl jeden z těchto nových kvalifikátorů také v souboru německém (DC.Abstract.Note).
8. Větší neshody lze z obr. 5/1 vyčíst u prvku „**Datum**“. U všech tří souborů je sice tento prvek povinný, uplatňuje se i stejné schéma pro zápis hodnot (norma ISO 8601 nebo komplexní schéma W3CDTF, které normu ISO 8601 zahrnuje), v hodnotách se však vyskytují rozličné typy dat. Australský soubor požaduje jediné datum - datum předložení disertace (s prohlášením o splnění všech předepsaných požadavků k přidělení hodnoty). Využívá k tomu v plné schodě se souborem DC už jen nekvalifikovaný prvek DC.date. Soubor ETD-MS definuje pouze nekvalifikovaný prvek a požaduje uvedení data z titulní stránky (většinou jde o datum obhajoby; datum předložení práce po případných opravách registruje automaticky systém). Detailní data ze „životního cyklu“ e-disertace umožňuje zapisovat německý soubor. Využívá k tomu pouze kvalifikované prvky (jeden je v souladu s DC, dva nikoliv). Jde 1) o povinně uváděné datum prohlášení kandidáta za doktora, ve

většině případů jde zároveň o datum ústní zkoušky (DC.Date.Accepted), 2) o nepovinně uváděné datum předložení práce k oponentskému řízení (DC.Date.Submitted) a 3) o rovněž nepovinně uváděné datum odevzdání povinného výtisku, resp. povinného elektronického souboru práce příslušné univerzitní knihovně (DC.Date.Created - platný kvalifikátor DC). Německý soubor navíc zahrnuje ještě jedno datum (povinné), a sice datum dodání metadat národní bibliografické centrále ke zpracování (DC.Date.Delivery), které není rovněž definováno v souboru DC.

9. Poslední komentář se týká prvků věnovaných údajům o „**Typu**“ kvalifikační práce, o elektronickém „**Formátu**“ dokumentu a specifickým údajům o „**Pokrytí**“ a o „**Zdroji**“. Australský soubor v současné době zcela vylučuje všechny tyto údaje. U prvních dvou uvedených odborníci ADT argumentují tím, že v rámci iniciativy DCMI stále nejsou k dispozici ustálené a úplné nabídky typů a podtypů a také formátů informačních zdrojů. Soubor ETD-MS definuje prvek typu práce jako povinný a prvek formátu jako nepovinný. Prozatím doporučuje zapisovat hodnoty ze slovníků nabízených iniciativou DCMI, ale jak bylo již výše uvedeno, je celý soubor metadat doplněn také o nový prvek „thesis.degree“, v jehož rámci lze přesně specifikovat informace týkající se ve značné míře také typu předkládané kvalifikační práce. Německý soubor uvádí typ práce i formát dokumentu jako povinný a podobně jako soubor ETD-MS doporučuje využívání seznamů hodnot nabízených iniciativou DCMI či spolupracujícími komunitami nebo hodnot vlastních (viz například hodnota Text.Habilitation). Soubor metadat ETD-MS jako jediný umožňuje nepovinně zápis specifické informace o časovém a geografickém vymezení obsahu práce (dc.coverage), který je využíván zejména pro práce z oblasti geografické a astronomické. Soubor německý pak jako jediný využívá ze specifikace DC prvku „Zdroj“ pro zápis údajů o původním zdroji online verze kvalifikační práce („DC.Source“).

Výše uvedené závěry z analýzy tří vybraných metadatových souborů signalizovaly v roce 2003, že existuje určité základní jádro metadat (dané v základu sémantikou DC), které dané komunity považují za potřebné k základnímu popisu elektronických vysokoškolských kvalifikačních prací. Odhlížíme při tom od drobných formálních rozdílů, které existují zejména na úrovni kvalifikátorů. Je nutné vzít v potaz, že analyzované soubory nejsou uzavřené, naopak počítá se s jejich dalším dopracováváním či zpřesňováním. Je zřejmé, že rozsáhlejší specifikace se budou objevovat u souborů definovaných na národních úrovních, zejména pokud v projektech participují národní bibliografické agentury (SRN ale i další země). Základní jádro bude pravděpodobně dále dotvářeno a zpřesňováno v rámci nejrozsáhlejšího systému NDLTD i v souvislosti s tím, jak se k němu budou hlásit další instituce z celého světa. Lokální systémy e-disertací, zakládané v současné době již také v ČR, si mohou v tuto chvíli vybrat některou z existujících zahraničních specifikací. Pokud se časem podaří případně připravit národní program, lze předpokládat, že se převezme již nějaká revidovaná a ustálená specifikace, pomocí níž bude možná nejenom bezproblémová národní ale i mezinárodní komunikace českých e-disertací.

Specifikace metadat systému ADT se v současnému datu změnila jen minimálně, specifikace ETD-MS také. Proměnou prošla specifikace METADISS, celonárodní německý systém DissonLine.de dnes používá specifikaci novou založenou na XML XMetaDiss [XMetaDiss, 2006]. Je založena již také na nejnovější specifikaci Dublin Core z roku 2008, je hluboce propracovaná (velký objem prvků popisných a administrativních včetně technických) a vyznačuje se novou koncepcí: zahrnuje prvky pro popis a komunikaci samotných prací a také prvky pro popis fyzických osob a korporací. Obr. 5/2 ukazuje aktuální metadatový záznam německé disertace ve formátu XMetaDiss (oai_xmetadiss) vyexportovaný pomocí protokolu OAI-PMH z repozitáře Humboldtovy univerzity (<http://edoc.hu-berlin.de/OAI-2.0>). Záznam byl získán přes program Repository Explorer (<http://re.cs.uct.ac.za/>).

<http://re.cs.uct.ac.za/>

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2009-03-28T14:44:49Z</responseDate>
  <request verb="GetRecord" metadataPrefix="oai_xmetadiss"
    identifier="oai:HUBerlin.de:29640">http://edoc.hu-berlin.de/OAI-2.0</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:HUBerlin.de:29640</identifier>
        <timestamp>2009-03-06</timestamp>
        <setSpec>pub-type:dissertation</setSpec>
        <setSpec>dnb:30</setSpec>
        <setSpec>ddc:540</setSpec>
      </header>
      <metadata>
        <xMetaDiss:xMetaDiss xmlns:xMetaDiss="http://www.d-
          nb.de/standards/xMetaDiss/"
          xmlns:dc="http://purl.org/dc/elements/1.1/"
          xmlns:dcterms="http://purl.org/dc/terms/"
          xmlns:ddb="http://www.d-nb.de/standards/ddb/"
          xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
          xmlns:pc="http://www.d-nb.de/standards/pc/"
          xmlns:thesis="http://www.ndltd.org/standards/metadata/etdms/
            1.0/" xmlns:cc="http://www.d-nb.de/standards/cc/"
          xmlns:urn="http://www.d-nb.de/standards/urn/"
          xsi:schemaLocation="http://www.d-nb.de/standards/xMetaDiss/
            http://www.d-nb.de/standards/xmetadiss/xmetadiss.xsd">
          <dc:title xsi:type="ddb:titleISO639-2" lang="eng">Functionality
            of noble-metal clusters</dc:title>
          <dc:title xsi:type="ddb:titleISO639-2" lang="eng">reactivity
            and optical properties</dc:title>
          <dc:creator xsi:type="pc:MetaPers">
            <pc:person>
              <pc:name type="nameUsedByThePerson">
                <pc:foreName>Christian</pc:foreName>
                <pc:surName>Bürgerl</pc:surName>
              </pc:name>
              <pc:dateOfBirth xsi:type="dcterms:W3CDTF">1979-
                03-06</pc:dateOfBirth>
              <pc:nativePlace
                Scheme="GKD">Berlin</pc:nativePlace>
            </pc:person>
          </dc:creator>
          <dc:subject xsi:type="xMetaDiss:DDC-SG">540</dc:subject>
          <dc:subject xsi:type="xMetaDiss:noScheme">optical
            properties</dc:subject>
          <dc:subject xsi:type="xMetaDiss:noScheme">Edelmetall-
            Cluster</dc:subject>
          <dc:subject xsi:type="xMetaDiss:noScheme">Chemische
            Reaktivität</dc:subject>
          <dc:subject xsi:type="xMetaDiss:noScheme">Optische
            Eigenschaften</dc:subject>
          <dc:subject xsi:type="xMetaDiss:noScheme">Ab initio
            Quantenchemie</dc:subject>
          <dc:subject xsi:type="xMetaDiss:noScheme">noble-metal
            clusters</dc:subject>
          <dc:subject xsi:type="xMetaDiss:noScheme">chemical
            reactivity</dc:subject>
          <dc:subject xsi:type="xMetaDiss:noScheme">a b initio quantum
            chemistry</dc:subject>
          <dcterms:abstract xsi:type="ddb:contentISO639-2"
            lang="ger">In dieser Dissertation wurden die
            ungewöhnlichen und einzigartigen Eigenschaften von
            Edelmetall-Clustern untersucht, die durch Quantum-
```

Confinement im Sub-Nanometer-Bereich entstehen. Dabei zeigt sich, dass die chemischen und physikalischen Eigenschaften und damit die Funktionen nicht vom Festkörper abgeleitet werden können und stark von der Anzahl der Atome abhängen. Die erzielten theoretischen Ergebnisse wurden in enger Zusammenarbeit mit experimentell arbeitenden Partnergruppen erzielt. Dabei hat sich gezeigt, dass durch die enge Kooperation zwischen Theorie und Experiment ein tiefes Verständnis von fundamentalen Prozessen und den zugrunde liegenden Mechanismen erlangt werden kann. Im Rahmen dieser Dissertation wurden die Reaktivität von geladenen Goldoxid-Clustern in der Gasphase, die ultraschnelle Dynamik von Edelmetall-Clustern und deren Komplexen sowie die optischen Eigenschaften von kleinen, deponierten Silber-Clustern untersucht und damit Beiträge geliefert, die einzigartigen Eigenschaften von Edelmetall-Clustern im Zusammenhang mit der heterogenen Katalyse und Nano-Optik besser zu verstehen.

```
</dcterms:abstract>
<dcterms:abstract xsi:type="ddb:contentISO639-2"
  lang="eng">In this thesis, the unique novel properties of
  noble metal clusters which arise in the sub-nanometer size
  regime due to quantum confinement have been theoretically
  explored. It has been demonstrated that by adding or
  removing a single atom the chemical and physical properties
  and functionality of noble-metal clusters can strongly change.
  The theoretical results have been derived in close
  cooperation with experimental findings of partner groups
  demonstrating that by joint theoretical and experimental
  efforts thorough understanding of fundamental processes
  and underlying mechanisms can be achieved. This thesis
  addresses the reactivity of charged gas-phase gold-oxide
  clusters in the context of the heterogeneous gold nano-
  catalysis, the ultrafast dynamical properties of noble-metal
  clusters and their complexes, and the optical properties of
  silver clusters at surfaces.</dcterms:abstract>
<dc:publisher xsi:type="cc:Publisher"
  type="dcterms:ISO3166">
  <cc:universityOrInstitution>
    <cc:name>Mathematisch-Naturwissenschaftliche
      Fakultät I</cc:name>
    <cc:place>Berlin</cc:place>
  </cc:universityOrInstitution>
  <cc:address>Unter den Linden 6, 10099
    Berlin</cc:address>
</dc:publisher>
<dc:contributor thesis:role="referee"
  xsi:type="pc:Contributor">
  <pc:person>
    <pc:name type="nameUsedByThePerson">
      <pc:foreName>Vlasta</pc:foreName>
      <pc:surName>Bonacic-Koutecký</pc:surName>
    </pc:name>
  </pc:person>
</dc:contributor>
<dc:contributor thesis:role="referee"
  xsi:type="pc:Contributor">
  <pc:person>
    <pc:name type="nameUsedByThePerson">
      <pc:foreName>Klaus</pc:foreName>
      <pc:surName>Rademann</pc:surName>
    </pc:name>
  </pc:person>
</dc:contributor>
<dc:contributor thesis:role="referee"
  xsi:type="pc:Contributor">
  <pc:person>
    <pc:name type="nameUsedByThePerson">
      <pc:foreName>Piercarlo</pc:foreName>
```



```

        <pc:surName>Fantucci</pc:surName>
      </pc:name>
    </pc:person>
  </dc:contributor>
  <dc:terms:dateAccepted xsi:type="dcterms:W3CDTF">2009-02-
    20</dc:terms:dateAccepted>
  <dc:terms:issued xsi:type="dcterms:W3CDTF">2009-03-
    04</dc:terms:issued>
  <dc:type
    xsi:type="ddb:PublType">ElectronicThesisandDissertation</
    dc:type>
  <dc:identifier xsi:type="urn:nbn">urn:nbn:de:kobv:11-
    10096477</dc:identifier>
  <dc:terms:medium
    xsi:type="dcterms:IMT">application/pdf</dc:terms:medium
  >
  <dc:language xsi:type="dcterms:ISO639-
    2">eng</dc:language>
  <dc:rights xsi:type="ddb:noScheme" />
  <dc:terms:accessRights xsi:type="ddb:access"
    ddb:type="dcterms:URI"
    ddb:kind="domain">http://edoc.hu-
    berlin.de/e_autoren/index-
    oa.php#copyrights</dc:terms:accessRights>
  = <thesis:degree>
    <thesis:level>thesis.doctoral</thesis:level>
    = <thesis:grantor xsi:type="cc:Corporate">
      = <cc:universityOrInstitution>
        <cc:name>Humboldt Universität zu
          Berlin</cc:name>
        <cc:place>Berlin</cc:place>
        </cc:universityOrInstitution>
      </thesis:grantor>
    </thesis:degree>
    <ddb:contact ddb:contactID="L6000-0473" />
    <ddb:fileNumber>1</ddb:fileNumber>
    <ddb:fileProperties ddb:fileName="buergel.pdf"
      ddb:fileID="file1" />
    <ddb:identifier ddb:type="URL">http://edoc.hu-
      berlin.de/dissertationen/buergel-christian-2009-02-
      20/PDF/buergel.pdf</ddb:identifier>
    <ddb:rights ddb:kind="free" />
  </xMetaDiss:xMetaDiss>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

```

Obr. 5/2: metadatový záznam disertace z archivu Humboldtovy univerzity ve struktuře XML exportovaného pomocí protokolu OAI-PMH pomocí příkazu GetRecord (ve formátu xmetadiss) [převzato 2009-03-25]

5.3 Návrh vývojové verze národního standardu EVSKP-MS

5.3.1 Úvod

Jedním z úkolů, který začala Odborná komise pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací AKVŠ (<http://www.evskp.cz/>) ihned po svém založení na jaře 2004 řešit, bylo vytvoření návrhu **národního metadatového standardu pro popis elektronických vysokoškolských kvalifikačních prací (eVŠKP)**. Důvodem byla skutečnost, že v té době již některé vysoké školy v ČR zahájily provoz nebo připravovaly projekt lokálního systému pro zpřístupňování eVŠKP, v jehož rámci byly připraveny buď vlastní lokálně platné specifikace metadat, nebo byly adaptovány specifikace z jiných systémů

(domácích nebo zahraničních). Velká variabilita lokálních specifikací je velkou překážkou pro budování plánovaného **národního registru eVŠKP** a popřípadě i **národního elektronického archivu úplných textů eVŠKP**, a proto bylo rozhodnuto připravit návrh tohoto standardu co nejdříve. Záměrem bylo, aby se národní standard stal pro lokální specifikace závazný, protože jen tak bude možná bezproblémová směna metadat mezi systémy a bude možná zároveň i integrace metadat o českých eVŠKP do mezinárodních systémů.

Přípravou standardu byla pověřena Pracovní skupina pro metadata. Oficiální práce na metadatovém standardu byly zahájeny v dubnu 2004. Základním východiskem návrhu standardu se staly výsledky podrobné analýzy významných zahraničních metadatových specifikací, která proběhla již v roce 2003 v rámci průzkumu tehdejšího stavu zahraničních systémů pro zpřístupňování e-VŠKP (součást jedné z grantových úloh FRVŠ [BRATKOVÁ, 2004a, část 3.2.3]). Výsledek analýzy vedl, kromě jiných faktorů, k rozhodnutí, že zpracování českého národního standardu bude v obecné rovině vázáno na specifikaci známého *Souboru metadatových prvků Dublin Core* [Dublin Core Metadata Element, 2004] a že v rovině specifické se v zájmu interoperability bude přihlížet ke specifikaci amerického souboru metadat pro elektronické disertační a diplomové práce *ETD-MS* [ETD-MS, 2008], který využívají americké i další univerzity sdružené v mezinárodním systému NDLTD, a dále, zejména pokud jde o kvalifikující informace, ke specifikaci velmi propracovaného německého národního standardu pro vysokoškolské práce *METADISS* [METADISS, 2003], který je společným formátem Německé knihovny a národního projektu SRN „Dissertationen Online“. Při tvorbě návrhu českého standardu sehrály důležitou roli také domácí zkušenosti s popisem těchto dokumentů ve vysokoškolských knihovnách.

V období od dubna 2004 do září 2005 vzniklo několik interních verzí návrhu standardu, které byly podrobeny připomínkování všech členů celé Odborné komise. Návrh *Metadatového souboru pro elektronické vysokoškolské kvalifikační práce (EVSKP-MS)* byl ve své verzi 0.1 zveřejněn na webových stránkách Odborné komise 7. listopadu 2005 [EVSKP-MS, 2005-11-07] – viz Příloha č. 4. Zainteresované komunity byly vyzvány k veřejnému připomínkování. Po vyhodnocení všech připomínek bude připravena definitivní 1. verze tohoto standardu. Návrh zahrnuje ucelenou **specifikaci jednotlivých metadatových prvků** pro popis eVŠKP obhájených na vysokých školách v ČR a základní vymezení metod **reprezentace zařazených prvků** pro potřeby jejich komunikace a sdílení v prostředí počítačových sítí.

5.3.2 Specifikace metadat

Návrh národního metadatového souboru *EVSKP-MS* zahrnuje definice celkem **22 prvků**. Povinné jádro tvoří polovina z nich, tj. 11 prvků. V následujícím přehledu jsou povinné prvky uvedeny s podtržením, nepovinné jsou uvedeny kurzivou a v kulaté závorce je uvedena opakovatelnost prvku:

1. Název VŠKP (Neopakovatelný)
2. *Podnázev VŠKP* (Opakovatelný)
3. *Překlad názvu VŠKP* (Opakovatelný)
4. *Překlad podnázvu VŠKP* (Opakovatelný)
5. Autor VŠKP (Neopakovatelný)
6. *Datum narození autora VŠKP* (Neopakovatelný)
7. *Předmětový popis VŠKP* (Opakovatelný)
8. Abstrakt VŠKP (Opakovatelný)
9. *Instituce archivující a zpřístupňující VŠKP* (Opakovatelný)
10. *Osoby podílející se na vedení VŠKP* (Opakovatelný)
11. *Osoby podílející se na oponování VŠKP* (Opakovatelný)

12. Datum vytvoření / odevzdání VŠKP (Neopakovatelný)
13. Datum obhájení VŠKP (Neopakovatelný)
14. Typ VŠKP (Opakovatelný)
15. Formát VŠKP (Opakovatelný)
16. Identifikátor VŠKP (Opakovatelný)
17. Jazyk VŠKP (Opakovatelný)
18. Práva k využívání VŠKP (Opakovatelný)
19. Přidělování akademický titul (Neopakovatelný)
20. Typ studijního programu (Neopakovatelný)
21. Studijní program a studijní obor (Neopakovatelný)
22. Instituce přiděluje titul (Neopakovatelný)

Jednotlivé prvky jsou ve specifikaci [EVSKP-MS, 2005-11-07, kap. 2.1] uvedeny českým pojmenováním (v titulku), v tabulkovém přehledu je pak postupně prezentováno jejich označení a definované kvalifikátory. Následuje informace o opakovatelnosti prvku v jednom záznamu a informace o jeho závaznosti (povinnosti). Popis zahrnuje detailní informaci o prvku včetně jeho zápisu. Připojené příklady syntaxe jsou prozatím doplňkem specifikace.

K bezproblémovým prvkům, na kterých se členové Komise v zásadě shodli, patří základní popisné údaje: **Název** (povinný) a **Podnázev VŠKP**, **Překlad názvu** a **Překlad podnázvu**, **Autor** (povinný) a **Datum narození autora** (minimálně rok). Zvažována byla povinnost překladů názvů prací do cizích jazyků, zejména angličtiny, na národní úrovni. Prvky zůstávají nepovinné, ale jednoznačně je doporučen překlad názvu do angličtiny u disertačních prací. Perspektivně se počítá s propojováním jmen autorů na soubory autoritních záznamů.

V zásadě k bezproblémovým prvkům patří také **Abstrakt VŠKP** (povinný). Jeho uvádění v češtině bylo ustanoveno jednomyslně. Překlad abstraktu do angličtiny je doporučen jako povinný pouze u disertačních prací, u dalších typů VŠKP (bakalářských, diplomových nebo rigorózních) lze tuto povinnost stanovit na úrovni lokálních systémů v rámci vnitřních předpisů. Zřejmě největším problémem při tvorbě specifikace byl předmětový popis. V nejbližší budoucnosti nelze v ČR počítat s jednotným předmětovým pořadacím systémem pro potřeby eVŠKP. V rámci prvku **Předmětový popis VŠKP** je proto definován kvalifikátor, který bude specifikovat užitý pořadací systém. Předmětový popis je na národní úrovni nepovinnou záležitostí. Překlady předmětových termínů do cizích jazyků jsou jen doporučením pro lokální systémy.

Jako bezproblémové se jeví prvky určené pro zápis jmen vedoucích a oponentů VŠKP, jejich uvádění je však nepovinné, na některých pracovištích mohou nastat objektivní potíže s jejich získáváním. Jednoznačně byla přijata definice prvku **Instituce archivující a zpřístupňující VŠKP**. Složitě, a to i na mezinárodní úrovni, je definování prvku **Data vytvoření / Odevzdání VŠKP** (povinný) a **Data obhájení VŠKP**. Charakter data je nezbytné řešit pomocí kvalifikátorů. Problémem je také jejich obtížné získávání na některých školách. V další verzi metadatového standardu se u datací mohou vyskytnout změny.

V rámci standardu byla dále definována skupina 4 povinných formálních prvků: **Typ**, **Formát**, **Identifikátor** a **Jazyk VŠKP**. U většiny z nich se předpokládá využívání číselníků. U identifikátorů se prozatím předpokládá využívání adres URL, intenzivně se však v zahraničí i u nás pracuje na zavedení trvalých identifikátorů (URN aj.). Použitá URL adresa do lokálního systému by měla být pokud možno trvalá (perzistentní). Komise v dalších verzích standardu zváží formalizaci prvku **Práva k využívání VŠKP**, a to i v souvislosti s novelou Autorského zákona.

Navrhovaný metadatový standard zahrnuje také 4 specifické prvky, které nejsou obsaženy v obecné metadatové specifikaci Dublin Core: **Přidělovaný akademický titul** (povinný), **Typ studijního programu**, **Studijní program a studijní obor** (povinný) a **Instituce přidávající titul** (povinný). Prvky byly adaptovány z amerického souboru *ETD-MS*. Komise zvážila jejich význam pro potřeby online vyhledávání a stanovila většinu z nich jako prvky povinné, přestože se v současné době u některých jeví jako problém získávání příslušných hodnot. U většiny se předpokládá využívání číselníků.

5.3.3 Reprezentace metadatových prvků

Součástí návrhu národního metadatového souboru *EVSKP-MS* jsou příklady syntaxe zápisu metadat, které by měly být návodem pro správce jednotlivých lokálních systémů, jak zajistit export dat do připravovaného národního registru. Při návrhu se Komise snažila neomezovat jednotlivé lokální systémy definicí jen jednoho povinného standardu zápisu a tak podobně jako standard Dublin Core [Dublin Core, 2003c] nabízí hned několik možných syntaxí zápisu. Jedná se konkrétně o následující syntaxe:

- RDF/XML
- XML
- HTML/XHTML

Národní registr metadat bude podporovat při importu všechna výše uvedená doporučení a umožní vzájemný převod mezi jednotlivými způsoby reprezentace metadat. Interní reprezentace metadat v lokálních systémech není nijak předepisována a je plně na tvůrčích každého jednotlivého systému.

Detailní popis kódování metadat je popsán v návrhu národního metadatového souboru *EVSKP-MS* [EVSKP-MS, 2005, kap. 2.2].

5.4 Národní metadatový standard *EVSKP-MS*, v. 1.1. pro popis VŠKP a standardy související (aktuální verze sta

5.4.1 Úvod

*Metadatový soubor pro popis elektronických vysokoškolských kvalifikačních prací v ČR (EVSKP-MS, verze 1.1 – viz Příloha č. 1) a související specifikace pro popis fyzických osob (PersCZ, verze 1.0 – viz Příloha č. 2) a pro popis korporací (CorpCZ, verze 1.0 – viz Příloha č. 3) byly připraveny v období únor - červenec 2008 jako národní metadatové standardy pracovní skupinou Odborné komise pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací AKVŠ (<http://www.evskp.cz/>). Návrhy standardů byly podrobeny připomínkování členů Odborné komise. Práce na uvedených standardech byly dohodnutou součástí Centralizovaného projektu MŠMT C1/2008 s názvem „Národní registr VŠKP a systém odhalování plagiátů“ (NR VŠKP), jehož hlavním řešitelem je Masarykova univerzita v Brně. Hlavní metadatový standard byl zveřejněn nejprve ve verzi 1.0 v květnu 2008 [EVSKP-MS, 2008-05-05] a po nezbytném krátkém testování byl doplněn a zveřejněn v jeho další verzi 1.1 [EVSKP-MS, 2008-07-15] na serveru Odborné komise v červenci 2008. Doplnkové související standardy byly zveřejněny ve svých verzích 1.0 již v květnu 2008 [PersCZ, 2008; CorpCZ, 2008]. Standardy byly po dalších testech přijaty a schváleny hlavním řešitelem Centralizovaného projektu jako **importovací formáty** (vedle jiných importovacích formátů) pro systém *Národního registru VŠKP (NR VŠKP)*.*

Import metadat o VŠKP z lokálních repozitářů do databáze *Národního registru VŠKP* však nebyl jediným cílem či záměrem vzniku uvedených metadatových standardů. Na úrovni

vstupů dat do lokálních repozitářů VŠKP jednotlivých škol mohou sloužit k **vlastnímu popisu** vysokoškolských kvalifikačních prací a popřípadě i k popisu fyzických osob a korporací, které se vznikem a komunikací prací souvisejí. Novinkou hlavního národního standardu je také vybraná množina administrativních a technických metadat určených k **přenosu úplných textů elektronických VŠKP** (směrem do *NR VŠKP*). Z obecného pohledu však byly uvedené standardy připraveny zejména jako **výměnné (komunikativní) formáty** (nově ve struktuře **XML** a kódování **UTF-8**), které mohou sloužit jak na úrovni importu, tak exportu mezi libovolnými informačními systémy. Předpokládají se toky metadat nejen směrem do *NR VŠKP* (<http://www.theses.cz/>), ale i z *NR VŠKP* směrem do jiných systémů či služeb. Počítá se s přenosem metadat do/z budoucího systému *NUŠL* (<http://nysl.stk.cz/>), počítá se s přenosem metadat do zahraničních registračních systémů VŠKP, jako je renomovaný mezinárodní systém *NDLTD* (<http://www.ndltd.org/>), evropský portál pro VŠKP *DART-Europe E-theses Portal (DEEP)* (<http://www.dart-europe.eu/>), nový evropský projekt komplexního systému pro šedou literaturu a informace související *DRIVER II* (<http://www.driver-community.eu/>) aj.

Jedním ze **základních východisek** přípravy uvedených standardů byla vývojová verze (0.1) specifikace *EVSKP-MS* z roku 2005 [*EVSKP-MS*, 2005], která sloužila od doby jejího zveřejnění vysokým školám pro potřeby popisu a komunikace metadat o VŠKP v jejich lokálních prostředích. Vývojová verze byla řešiteli Centralizovaného projektu *NR VŠKP* také přijata jako možný importovací formát. I proto se zpracovatelé nové verze (1.0 a 1.1) specifikace *EVSKP-MS* snažili zachovat původní jádro povinných prvků tak, aby přechod na novou verzi byl pokud možno bezproblémový. Vývojová verze byla připravena v souladu s mezinárodními standardy platnými v dané době, které byly v roce 2008 aktualizovány. Vývojová verze neobsahovala informace potřebné pro přenos a zpřístupnění plných textů.

Základní skupina popisných prvků *EVSKP-MS* vychází, stejně jako jiné ve světě známé národní metadatové specifikace, především z aktuální verze 1.1 (ze 14. ledna 2008) doporučení obecného standardu pro popis informačních zdrojů *Dublin Core* [Dublin Core Metadata Element, 2008], resp. z aktuální verze definic *Termínů DCMI* rovněž ze 14. ledna 2008 [DCMI Metadata Terms, 2008]. Specifikace *Dublin Core* doznala od doby zveřejnění nulté verze standardu *EVSKP-MS* značných proměn. Nejdříve v roce 2006 a následně i v roce 2008. V důsledku toho se první verze standardu *EVSKP-MS* (1.0) musela oproti verzi předchozí v mnoha směrech změnit.

Některé specifické prvky pro popis vysokoškolských kvalifikačních prací byly převzaty z aktuální verze 1.00 (rev. 2) nejvýznamnějšího mezinárodního standardu *ETD-MS* v z roku 2008 [*ETD-MS*, 2008], který byl připraven pro potřeby mezinárodního decentralizovaného systému *NDLTD*. Při tvorbě českého národního standardu bylo dále přihlédnuto z hlediska koncepčního i z hledisek jiných k vybraným národním standardům. Jde zejména o nejnovější verzi 1.3 z roku 2006 německého národního metadatového standardu *XMetaDiss* pro detailní popis disertací [*XMetaDiss*, 2006]. Německý standard je připraven na zcela nové bázi (též v souladu se změnami specifikace *Dublin Core*), předpokládá komunikaci metadat již jen ve struktuře XML a u vybraných základních popisných prvků je nabízeno vnořování prvků popisujících fyzické osoby a korporace. Přihlédnuto bylo také ke starší německé specifikaci metadat autoritních entit (fyzických osob) *MetaPers* ve verzi 1.1 z roku 2003 [*MetaPers*, 2003]. Konzultována byla také aktuální 2. verze návrhu francouzské normy specifikující metadata pro elektronické disertace *TEF 2.0* z března 2006 [*TEF 2.0*, 2006], zejména pak definice některých prvků (například definice a atributy prvku `dc:language` aj.). Detailní popisy prvků českého národního standardu přihlížejí k praxi popisu VŠKP v ČR.

Nově byla přijata **koncepce tří souběžných standardů**: pro popis prací a pro popis fyzických osob a korporací. Prvky pro popis fyzických osob a korporací mohou být volitelně použity

jako **prvky vnořené** u vybraných prvků popisu VŠKP, ale i v dalších standardech. Základní popisné prvky *EVSKP-MS* (a jejich povinnost uvádění) byly v zásadě zachovány, přibýly však některé prvky jiné (jako **nepovinné**). Větší změny nastaly v oblasti kvalifikujících informací, došlo k **zpřesňování prvků** a také ke **zpřesňování jejich atributů**. V souladu se standardem *Dublin Core* byla ve jménech prvků zavedena **dvojtečka** místo původní tečky. Dvojtečka odděluje jmenný prostor od další části jména prvku (dc: , dcterms: , thesis: , evskp:). Další část jména prvku již není vnitřně dělena žádnou interpunkcí (zrušeno bylo vícenásobné uvádění teček v nulté verzi specifikace). Pro odlišení více slov ve jménu prvku je v souladu se standardem *Dublin Core* použito tzv. „camelCase“ formátování (například: dcterms:dateSubmitted, evskp:fileNumber aj.)

Standard *EVSKP-MS* zahrnuje v nové verzi také sadu prvků technického a administrativního charakteru. Jejich přítomnost je nezbytná pro zajištění přenosu samotných souborů s plnými texty mezi registry.

5.4.2 Popisné, technické a administrativní prvky EVSKP-MS

Hlavní národní metadatový soubor *EVSKP-MS*, verze 1.1 zahrnuje definice **1 hlavního kořenového prvku** (evskp:metadata) a **24 popisných a 8 technických a administrativních prvků**. **Povinné jádro tvoří 13 popisných metadatových prvků** (kromě prvku kořenového). Všechny technické a administrativní prvky jsou nepovinné. V následujícím přehledu jsou povinné prvky uvedeny s podtržením, nepovinné jsou uvedeny *kurzivou*, v kulaté závorce je uvedena opakovatelnost prvku:

Popisné metadatové prvky:

- evskp:metadata (Neopakovatelný)
 - dc:title (Opakovatelný)
 - *dc:terms:alternative* (Opakovatelný)
 - dc:creator (Neopakovatelný)
 - *dc:subject* (Opakovatelný)
 - dcterms:abstract (Opakovatelný)
 - *dcterms:tableOfContents* (Opakovatelný)
 - *dc:publisher* (Opakovatelný)
 - *dc:contributor* (Opakovatelný)
 - *dcterms:created* (Neopakovatelný)
 - *dcterms:dateSubmitted* (Neopakovatelný)
 - dcterms:dateAccepted (Neopakovatelný)
 - *dcterms:modified* (Opakovatelný)
 - dc:type (Opakovatelný)
 - dcterms:medium (Opakovatelný)
 - *dcterms:extent* (Opakovatelný)
 - dc:identifier (Opakovatelný)
 - dc:language (Opakovatelný)
 - *dcterms:bibliographicCitation* (Opakovatelný)
 - *dc:rights* (Opakovatelný)
 - thesis:degree (Neopakovatelný)
 - thesis:name (Neopakovatelný)

- thesis:level (Opakovatelný)
- thesis:discipline (Neopakovatelný)
- thesis:grantor (Neopakovatelný)

Technické a administrativní metadatové prvky

- *evskp:contact* (Neopakovatelný)
- *evskp:fileNumber* (Neopakovatelný)
- *evskp:fileProperties* (Opakovatelný)
- *evskp:transfer* (Opakovatelný)
- *evskp:server* (Neopakovatelný)
- *evskp:dateDelivered* (Neopakovatelný)
- *dcterms:available* (Neopakovatelný)
- *evskp:modified* (Opakovatelný)

Jednotlivé prvky jsou ve specifikaci [EVSKP-MS, 2008-07-15, část 2.1] prezentovány v tabulkové formě. Jsou uvedeny záhlavím se jménem prvku (tučně). V tabulce je ve stanoveném pořadí uváděno:

- **Označení** prvku (pojmenování v českém jazyce)
- **Jméno** prvku (odpovídá stanovené syntaxi zápisu v XML)
- **URI** (jednoznačný trvalý identifikátor prvku)
- **URI nadřazeného prvku** (jednoznačný trvalý identifikátor přímo nadřazeného prvku, v jehož rámci je uvedený prvek vnořen; pokud údaj není uveden, prvek není podřazen žádnému prvku mimo prvek kořenový)
- **Povinnost** prvku (ano/ne)
- **Opakovatelnost** prvku (ano/ne)
- **Atribut povinný** (zpřesňuje daný prvek)
- **Atribut volitelný** (zpřesňuje, pokud je zadán, daný prvek)
- **Definice** prvku (textový popis specifikuje jeho použití v metadatovém záznamu)
- **Poznámky** (doplňující informace určená ke zpřesnění popisu prvku a jeho použití)
- **Příklady** (ukázky použití prvku v zápisu XML)

5.4.2.1 Popisné prvky EVSKP-MS

K bezproblémovým prvkům nejnovější verze specifikace *EVSKP-MS* patří základní popisné údaje. Jsou to především již i dříve zahrnuté prvky **Název VŠKP** (dc:title, povinný) a nepovinný prvek **Podnázev VŠKP** (dcterms:alternative). Povinně uváděný atribut (xml:lang) specifikuje pomocí kódu normy ISO 639 jazyk názvových informací. Nově jsou řešeny údaje o překladu názvu a podnázvu. V nulté verzi byly samostatnými prvky, ve verzi nové je lze zapisovat pouze v rámci výše uvedených prvků pro název a podnázev VŠKP pomocí volitelného atributu (evskp:typeTranslated) s hodnotou atributu "translated". Jde o požadavek nové specifikace standardu *Dublin Core* [Dublin Core Metadata Element, 2008]. Překlady názvů jsou u běžných VŠKP nepovinné, jednoznačně je ale doporučen překlad názvu do angličtiny u disertačních prací. Prvek **Autor** (dc:creator, povinný) je rovněž zachován s tím, že nová specifikace umožňuje jak zápis původní jednoduché formy jména autora s patřičnou syntaxí, tak zápis nové, strukturované formy jména autora (s vnitřním

vnořováním specifických prvků charakterizujících autora) na základě související samostatné specifikace *PersCZ* (v prvku *pcz:person*) [PersCZ, 2008]. Ve vývojové verzi zařazený samostatný prvek Datum narození autora je nyní možné uvádět v rámci strukturované formy zápisu jména autora jako samostatný prvek (*dateOfBirth*). Možné je méně preferované uvedení data narození jako přídatné informace v rámci jednoduché formy jména autora (po středníku a mezeře).

K bezproblémovým prvkům patří v zásadě také prvky věnované obsahu VŠKP. V nové specifikaci zůstává i nadále prvek **Abstrakt VŠKP** (*dcterms:abstract*, povinný). Jeho povinné uvádění v češtině je jednoznačné, zájemci však mohou zapisovat text abstraktu také v jiných jazycích (bude uveden povinný atribut *xml:lang*. Překlad abstraktu do angličtiny je povinný pouze u disertačních prací, u dalších typů VŠKP (bakalářských, diplomových nebo rigorózních) lze tuto povinnost stanovit na úrovni lokálních systémů v rámci vnitřních předpisů. Novým rozšiřujícím, ale nepovinným prvkem, je prvek **Obsah VŠKP** (*dcterms:tableOfContents*). Zájemci mohou v rámci popisu uvádět různě strukturovaný (i na více hierarchických úrovních) vlastní obsah VŠKP. Z určitého hlediska je v ČR stále největším problémem uvádění hodnot prvku **Věcný popis VŠKP** (*dc:subject*). V nejbližší budoucnosti nelze v ČR počítat s jednotným předmětovým nebo systematickým pořádacím systémem pro potřeby popisu VŠKP u všech zainteresovaných škol. V rámci prvku věcného popisu je proto nově definován nepovinný atribut (*evskp:typeSubject*), v jehož rámci je nutné v kódovaném tvaru specifikovat užitý pořádací systém (jejich možné aktuální seznamy jsou ve specifikaci odkazovány). Předmětový popis je na národní úrovni nepovinnou záležitostí. Překlady předmětových termínů do cizích jazyků jsou možné, využívá se opět povinného atributu *xml:lang*.

Jednoznačně je ve specifikaci (již z dřívější nulté verze) přijata definice prvku **Instituce archivující a nebo zpřístupňující VŠKP** (*dc:publisher*). Původní jednoduchá forma zápisu jména korporace či instituce s patřičnou syntaxí je zachována, nově je ale možné využít strukturované formy jména (s vnitřním vnořováním specifických prvků charakterizujících korporaci) na základě doplňkové specifikace *CorpCZ* (v prvku *universityOrInstitution*) [CorpCZ, 2008].

Formální změny nastaly u prvků, které dříve nesly jména vedoucích a oponentů VŠKP. Původní dva samostatné prvky byly (opět v souladu se zahraničními standardy) sloučeny do jediného prvku **Vedoucí nebo oponent VŠKP** (*dc:contributor*) s tím, že je zaveden jeho povinný atribut (*thesis:role*), v jehož rámci se pomocí daných hodnot ("advisor" nebo "referee") zapisuje konkrétní role fyzické osoby. Uvádění tohoto prvku je nepovinné, i když příslušné údaje mohou producenti dat získat z databáze vysoké školy (například studijního systému). Stejně jako u prvku Autor VŠKP je i u tohoto prvku možné zapisovat jméno fyzické osoby jednoduchým způsobem nebo pomocí nového standardu *PersCZ* [PersCZ, 2008].

Formální změny nastaly u datací spojených s VŠKP. Nová specifikace nabízí prvky pro zápis jednoznačněji specifikovaných dat. Povinným prvkem je **Datum obhajoby VŠKP** (*dcterms:dateAccepted*), který lze dnes automaticky přebírat ze školních studijních agend. Nepovinně lze v záznamu uvádět **Datum vytvoření VŠKP** (*dcterms:created*), **Datum odevzdání či podání VŠKP** (*dcterms:dateSubmitted*) a dokonce i **Datum modifikace VŠKP** (*dcterms:modified*), které se může uplatnit například v rámci migračních či archivačních procesů. Hodnoty dat se zapisují v souladu se standardem *ISO 8601*.

V rámci nejnovější verze standardu *EVSKP-MS* byly dále zachovány 4 povinné formální prvky (dva nepovinné přibyly). V rámci prvku **Typ VŠKP** (*dc:type*) se na prvním místě povinně uvádí typ VŠKP v českém jazyce na základě platného číselníku (TypVSKP) pomocí

povinného atributu (evskp:typeType). V nepovinných výskytech prvku lze zapisovat také hodnoty známé typologie informačních zdrojů DCMI (dcterms:DCMIType). Původní prvek Formát (z nulté verze) zůstává součástí nové specifikace, došlo však k jeho formálnímu zpřesnění na prvek **Médium (formát souboru) VŠKP** (dcterms:medium). Hodnoty formátu se zapisují podle autorizovaného seznamu organizace IANA. Nepovinným „přídavkem“ k prvku Médium v nové specifikaci je prvek **Rozsah VŠKP** (dcterms:extent). Umožňuje zápis podrobnějších údajů o rozsahu nebo velikosti VŠKP (například počet stran u papírových VŠKP, které byly digitalizovány). Prvek **Identifikátor** (dc:identifier) je podstatným povinným prvkem specifikace (stejně jako v nulté verzi), jde o zápis jednoznačného identifikátoru VŠKP. Zapisovat lze i URL s tím, že pro budoucnost je optimální nasazování některého z trvalých identifikátorů (URN, ARK aj.). Některé vysoké školy v ČR již trvalé identifikátory užívají, např. HDL (Handle) v rámci softwaru DSpace. Prvek **Jazyk VŠKP** (dc:language) zůstává i nadále součástí specifikace *EVSKP-MS* s tím, že zpracovatelé zvolili možnost 2- nebo 3- znakového kódu podle standardu *ISO 639*. Novým nepovinným prvkem je prvek **Bibliografická citace VŠKP** (dcterms:bibliographicCitation), který je určen zájemcům o uvedení citačního záznamu VŠKP podle standardů *ČSN ISO 690* a *ČSN ISO 690-2*.

V nové specifikaci zůstává také nepovinný prvek **Práva k využívání VŠKP** (dc:rights), jehož doplňkem na úrovni definice je uvedení možnosti odkazu URI na zdroj informací s popisem práv k využívání VŠKP (např. URL odkaz na licenční smlouvu).

Verze 1.1 národního standardu *EVSKP-MS*, stejně jako jeho nultá verze, zahrnuje popisné prvky převzaté z amerického standardu *ETD-MS* [ETD-MS, 2008], které nejsou obsaženy v obecné metadatové specifikaci *Dublin Core* [Dublin Core Metadata Element, 2008]. Jde celkem o 5 specifických prvků, které jsou formálně nově koncipované tak, že první z nich **Akademický titul nebo vědeckopedagogická hodnost** (thesis:degree) je v roli nadřazeného prvku, ostatní v roli prvků vnořených. Všechny prvky jsou povinné. Patří k nim prvek **Zkratka jména akademického titulu nebo vědecko-pedagogické hodnosti** (thesis:name) - příslušné hodnoty se zapisují v souladu se zněním *Zákona o VŠ/1998*, dále prvek **Typ studijního programu** (thesis:level) s doporučeným uváděním hodnot opět podle *Zákona o VŠ/1998*, prvek **Studijní program a studijní obor** (thesis:discipline), jehož hodnoty jsou zapisovány podle stanovených číselníků, a prvek **Instituce přidělující titul** (thesis:grantor). U posledního prvku nová verze standardu umožňuje dvojí způsob zápisu informací o instituci stejně, jako tomu je u prvku dc:publisher (též v souladu se specifikací *CorpCZ*).

5.4.2.2 Technické a administrativní prvky EVSKP-MS

Pro přenos metadatových záznamů a souborů VŠKP z registru lokálního do registrů dalších, např. do NR VŠKP, se ukázala potřeba přípravy skupiny nepovinných technických a administrativních prvků, které přímo nepopisují samotnou kvalifikační práci, ale **evidují změny v metadatech a informují o umístění a zpřístupnění souborů VŠKP**. Tyto specifické prvky patří, až na prvek dcterms:available, do nově definovaného jmenného prostoru evskp.

První z těchto prvků (evskp:contact) označuje školu nebo přímo fakultu, na které byla práce obhajována. **Resortní identifikátor školy nebo fakulty RID** eviduje celostátní matrika studentů a je uveden v databázi REGPRO Ústavu pro informace ve vzdělávání. RID se zapisuje jako atribut contactID prvku evskp:contact, jméno školy nebo fakulty může být zapsáno do obsahu prvku. Prvek je povinný pro identifikaci poskytovatele metadat v NR VŠKP.

Prvek evskp:server obsahuje **informace o instituci zodpovědné za poskytování metadat**. Do prvku evskp:server se zapisuje buď přímo jméno instituce, nebo je instituce detailněji popsána pomocí standardu *CorpCZ*.

Významnou skupinu technických a administrativních prvků tvoří prvky, jejichž obsahem je datum.

Pokud je metadatový záznam naimportován do dalšího úložiště, **datum importu** je uloženo v prvku evskp:dateDelivered.

Pokud dojde ke změně metadatového záznamu, např. k úpravě abstraktu nebo klíčových slov knihovníkem, **datum změny záznamu** eviduje prvek evskp:modified. Toto datum je důležité pro export metadat – pomocí data modifikace záznamu lze zjistit, které záznamy byly změněny od data posledního importu. Datum posledního importu si eviduje každý z registrů, který metadata importuje. V rámci popisných prvků existuje podobný prvek dcterms:modified, který ale eviduje datum změny samotného souboru VŠKP popsaného metadatovým záznamem.

V některých situacích je potřeba **omezit přístupnost plných textů** od určitého data nebo po daný časový interval. Ve standardu *EVSKP-MS* je možné toto omezení stanovit v prvku dcterms:available.

Pro **popis a přenos souborů** vznikly tři nové prvky. Prvek evskp:fileNumber uvádí číslem počet všech souborů, které tvoří kvalifikační práci, resp. které jsou v metadatovém záznamu odkazovány. Pro každý tento soubor v záznamu existuje prvek evskp:fileProperties, který v obsahu prvku uvádí text k zobrazení uživateli (např. název nebo popis obsahu souboru) a v attributech prvku popisné informace o souboru. Mezi tyto atributy (z většiny nepovinné) patří např. typ souboru (kvalifikační práce, oponentura, příloha apod.), velikost souboru, formát souboru nebo jeho kódová stránka. Tyto informace jsou důležité při rozhodování o stažení souboru či v případě potřeby dohledat specifickou skupinu souborů, např. při nutnosti migrace specifického formátu souborů na novější verzi.

Samotnou **URL adresu pro přenos souborů** VŠKP zahrnuje prvek evskp:transfer. Buď obsahuje odkaz na jeden ZIP soubor s archivem všech souborů VŠKP, nebo pro každý ze souborů existuje samostatný prvek evskp:transfer s URL adresou tohoto souboru. V případě ZIP archivu se jednotlivé soubory, popsané v prvcích evskp:fileProperties, identifikují adresářem a jménem prvku uvedených ve dvou attributech prvků evskp:fileProperties.

5.4.3 Popisné prvky standardů PersCZ a CorpCZ

K metadatovému standardu pro popis VŠKP byly připraveny dva související standardy, a to pro **popis fyzických osob** (např. autoři VŠKP, vedoucí a oponenti VŠKP aj.) *PersCZ*, verze 1.0 [PersCZ, 2008], a pro **popis korporací** (např. instituce archivující a zpřístupňující VŠKP, instituce přidělující tituly, afiliace aj.) *CorpCZ*, verze 1.0. [CorpCZ, 2008]. Zpracovatelé se rozhodli připravit tyto standardy z důvodu potřeby uchovávat popis fyzických osob a korporací platný k dané době. Pro tyto účely nelze vždy použít bázi autoritních dat. Užívání obou specifikací ve standardu EVSKP-MS je dobrovolné.

První specifikace zahrnuje v rámci povinného **kořenového prvku (person)** definice **13 popisných metadatových prvků**, které popisují vybrané vlastnosti fyzických osob. V následujícím přehledu jsou povinné prvky uvedeny s podtržením (jde o dva prvky), nepovinné jsou uvedeny *kurzívou* a v kulaté závorce je uvedena opakovatelnost prvku:

- person (Opakovatelný)
 - *dc:identifier (Opakovatelný)*

- *academicTitleBefore (Opakovatelný)*
- name (Neopakovatelný)
 - *foreName (Neopakovatelný)*
 - *surName (Neopakovatelný)*
 - *personEnteredUnderGivenName (Neopakovatelný)*
- *academicTitleAfter (Opakovatelný)*
- *dateOfBirth (Neopakovatelný)*
- *placeOfBirth (Neopakovatelný)*
- *note (Opakovatelný)*
- *email (Opakovatelný)*
- *homepage (Opakovatelný)*
- *affiliation (Opakovatelný)*

Základním povinným nadřazeným prvkem této specifikace je prvek **Popis fyzické osoby** (person, jmenný prostor pcz). K základním vnořeným prvkům patří zatím nepovinný prvek **Identifikátor fyzické osoby** (dc:identifier). Dosazování reálných hodnot bude závislé od vývoje systémů pro správu takových identifikátorů v ČR. Prozatím známe takové identifikátory zejména z některých zahraničních systémů (např. trvalý identifikátor ARK pro fyzické osoby, digitální identifikátor DAI z nizozemské praxe aj.). K nepovinným prvkům patří zejména prvky, které popisují vlastnosti fyzických osob: získané tituly před jménem a za jménem, datum narození, místo narození, emailová adresa, personální domovská stránka a afiliace. Povinným prvkem specifikace je prvek **Jméno** (name), v jehož rámci jsou detailně specifikovány na úrovni vnořených prvků buď jednotlivé části běžného strukturovaného jména, nebo výjimečně jméno nestrukturované.

Druhá specifikace zahrnuje v rámci povinného **kořenového prvku** definice **8 popisných metadatových prvků**. V následujícím přehledu jsou povinné prvky uvedeny s podtržením (jde také o dva prvky), nepovinné jsou uvedeny kurzivou a v kulaté závorce je uvedena opakovatelnost prvku:

- universityOrInstitution (Opakovatelný)
 - *dc:identifier (Opakovatelný)*
 - name (Opakovatelný)
 - *place (Opakovatelný)*
 - *address (Opakovatelný)*
 - *email (Opakovatelný)*
 - *homepage (Opakovatelný)*
 - *note (Neopakovatelný)*
 - *department (Neopakovatelný)*

Základním povinným kořenovým prvkem této specifikace je prvek **Popis univerzity nebo instituce** (universityOrInstitution, jmenný prostor ccz). K základním vnořeným prvkům patří i v tomto případě zatím nepovinný prvek **Identifikátor korporace** (dc:identifier). Dosazování reálných hodnot bude také závislé od vývoje systémů pro správu takovýchto identifikátorů v ČR. K nepovinným prvkům patří zejména prvky, které popisují vybrané vlastnosti korporací: místo sídla instituce, adresa sídla, emailová adresa nebo domovská stránka. Povinným prvkem specifikace je prvek **Jméno** (name) - na vnořené úrovni lze uvádět i jméno **Podřízené jednotky instituce** (department).

5.4.4 Formát zápisu metadatových prvků a validace záznamu

Při volbě formátu zápisu prvků se zpracovatelé standardu přiklonili k nejčastějšímu způsobu zápisu – tzv. XML „Vanilla“ kódování, které zmiňuje např. americký standard *ETD-MS* [ETD-MS, 2008]. Doporučené kódování znaků v XML záznamu je UTF-8.

Ve verzi 1.1 standardu *EVSKP-MS* jsou nově uvedena **schémata s definicí prvků**, která lze použít pro automatizovanou validaci konkrétního metadatového záznamu. Schémata vyžadují dodržení pořadí prvků podle standardu *EVSKP-MS*. Pro standardy *EVSKP-MS*, *CorpCZ* a *PersCZ* byla vypracována tato schémata:

- XML schéma
- RELAX NG schéma
- RELAX NG schéma (kompaktní syntaxe)

URL adresy schémat jsou součástí jednotlivých standardů. **Validace metadatových záznamů** lze, pomocí výše uvedených schémat, provést např. online na WWW (<http://validator.nu>).

Následující 2 příklady jsou z reálné praxe a dokladují platnost standardů *EVSKP-MS*. Na obr. 5/3 je záznam diplomové práce z repozitáře DSpace Akademie múzických umění v Praze (<http://dspace.amu.cz/oai/request>) ve formátu *EVSKP-MS* (oai_evskpms). Záznam nemá vnořované prvky pro fyzické osoby a korporace (jde o jednu z možností využití formátu *EVSKP-MS*). Na obr. 5/4 je záznam stejné diplomové práce – nyní ve struktuře *METS* (automatická konverze do tohoto formátu je zajišťována softwarem DSpace). Záznamy z AMU mají uplatněn trvalý identifikátor HDL (americký Handle), konkrétně (<http://hdl.handle.net/10318/148>).

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2009-03-27T15:31:36Z</responseDate>
  <request identifier="oai:dspace.amu.cz:10318/148" metadataPrefix="oai_evskpms"
    verb="GetRecord">http://dspace.amu.cz/oai/request</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:dspace.amu.cz:10318/148</identifier>
        <timestamp>2009-02-13T10:13:51Z</timestamp>
        <setSpec>hdl_10318_126</setSpec>
      </header>
      <metadata>
        <evskp:metadata version="1.1"
          xmlns:ccz="http://www.evskp.cz/standardy/corpcz/"
          xmlns:dc="http://purl.org/dc/elements/1.1/"
          xmlns:dcterms="http://purl.org/dc/terms/"
          xmlns:evskp="http://www.evskp.cz/standardy/evskp/"
          xmlns:pcz="http://www.evskp.cz/standardy/perscz/"
          xmlns:thesis="http://www.ndltd.org/standards/metadata/etdms/1.0/">
          <dc:title xml:lang="cs">Psychologické působení barev a jeho
            využití v barevné kinematografii</dc:title>
          <dcterms:alternative xml:lang="cs">Psychological Working of
            Colours and its Application in Colour
            Cinematography</dcterms:alternative>
          <dc:creator>Strnad, Jan</dc:creator>
          <dc:subject xml:lang="cs">Film</dc:subject>
          <dcterms:abstract xml:lang="en">Thesis trats of elementary
            physical color characteristics, principle of human color
            perception. Refer to psychosomatic influence of colors on
```


human personality. In more detail follows color as means of expression, and its usage in cinematography, color harmony, color composition etc. Psychologic, psychophysiologic and symbolic study of chosen colors is concluding part of thesis.

```

</dc:terms:abstract>
<dc:terms:abstract xml:lang="cs">Práce pojednává o základních fyzikálních vlastnostech barev, principu jejich vnímání z hlediska fyziologie vidění, poukazuje na psychosomatické vlivy jednotlivých barev na lidskou osobnost. Podrobněji sleduje barvy jako nástroje obrazové skladby, a jejich využití v kinematografii, barevnou harmonii a barevnou kompozici. Jejím závěrem je psychologický, psychofyziologický a symbolický rozbor některých vybraných barev.</dc:terms:abstract>
<dc:publisher xml:lang="cs">Knihovna Filmové fakulty Akademie múzických umění</dc:publisher>
<dc:contributor thesis:role="advisor">ŠOFŘ, Jaromír</dc:contributor>
<dc:contributor thesis:role="referee">BRABEC, Jaroslav</dc:contributor>
<dc:terms:created>2007</dc:terms:created>
<dc:terms:dateSubmitted>2007-09-24</dc:terms:dateSubmitted>
<dc:terms:dateAccepted>2007-10-10</dc:terms:dateAccepted>
<dc:type xml:lang="cs" evskp:typeType="TypVSKP">Diplomová práce</dc:type>
<dc:terms:medium>application/msword</dc:terms:medium>

<dc:identifier>http://hdl.handle.net/10318/148</dc:identif
ier>
<dc:language>cs</dc:language>
<dc:rights xml:lang="cs">a) rozmnožováním díla jakýmikoliv prostředky a v jakékoliv formě, včetně formy elektronické</dc:rights>
<dc:rights xml:lang="cs">b) bezúplatným rozšiřováním originálu Školního díla nebo jeho rozmnoženin pořízených podle písm. a)</dc:rights>
<dc:rights xml:lang="cs">d) vystavováním originálu díla nebo jeho rozmnoženin podle písm. a)</dc:rights>
<dc:rights xml:lang="cs">e) sdělováním veřejnosti všemi způsoby., zejména pak živým provozování díla....a zpřístupňováním v síti Internet...</dc:rights>
<thesis:degree>
<thesis:name>MgA.</thesis:name>
<thesis:level xml:lang="cs">Magisterský studijní program</thesis:level>
<thesis:discipline xml:lang="cs">Filmové, televizní a fotografické umění a nová média / Kamera</thesis:discipline>
<thesis:grantor xml:lang="cs">Akademie múzických umění v Praze. Filmová s televizní fakulta.</thesis:grantor>
</thesis:degree>
<evskp:contact contactID="S5131" />
<evskp:fileProperties fileName="AMU test.pdf" fileID="757" fileSize="14800" />
<evskp:fileProperties fileName="AMU test.pdf.txt" fileID="758" fileSize="58">Extracted text</evskp:fileProperties>
<evskp:transfer fileID="757" accessRights="domain">http://dspace.amu.cz/bitstream/10318/148/1/AMU%20test.pdf</evskp:transfer>
<evskp:transfer fileID="758" accessRights="domain">http://dspace.amu.cz/bitstream/10318/148/2/AMU%20test.pdf.txt</evskp:transfer>
<dc:terms:available>2008-11-02</dc:terms:available>
<evskp:modified>2009-02-13T10:13:51Z</evskp:modified>
</evskp:metadata>
</metadata>
</record>
</GetRecord>

```

</OAI-PMH>

Obr. 5/3: metadatový záznam diplomové práce z archivu AMU ve struktuře XML exportovaného pomocí protokolu OAI-PMH pomocí příkazu GetRecord (ve formátu oai_evskpms) [převzato 2009-03-25]

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2009-03-28T14:34:28Z</responseDate>
  <request identifier="oai:dspace.amu.cz:10318/148" metadataPrefix="mets"
    verb="GetRecord">http://dspace.amu.cz/oai/request</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:dspace.amu.cz:10318/148</identifier>
        <timestamp>2009-02-13T10:13:51Z</timestamp>
        <setSpec>hdl_10318_126</setSpec>
      </header>
      <metadata>
        <mets OBJID="hdl:10318/148" LABEL="DSpace Item"
          xmlns="http://www.loc.gov/METS/"
          xmlns:xlink="http://www.w3.org/1999/xlink"
          xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
          xmlns:mods="http://www.loc.gov/mods/v3"
          xsi:schemaLocation="http://www.loc.gov/METS/
            http://www.loc.gov/standards/mets/mets.xsd
            http://www.loc.gov/mods/v3
            http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
          <metsHdr CREATEDATE="2009-03-28T15:34:28">
            <agent ROLE="CUSTODIAN" TYPE="ORGANIZATION">
              <name>DSpace na AMU</name>
            </agent>
          </metsHdr>
          <dmdSec ID="DMD_hdl_10318/148">
            <mdWrap MDTYPE="MODS">
              <xmlData>
                <mods:name>
                  <mods:role>
                    <mods:roleTerm
                      type="text">advisor</mods:roleTerm>
                  </mods:role>
                  <mods:namePart>ŠOFR,
                    Jaromír</mods:namePart>
                </mods:name>
                <mods:name>
                  <mods:role>
                    <mods:roleTerm
                      type="text">author</mods:roleTerm>
                  </mods:role>
                  <mods:namePart>Strnad,
                    Jan</mods:namePart>
                </mods:name>
                <mods:extension>
                  <mods:dateAccessioned
                    encoding="iso8601">2008-11-
                      02T19:12:26Z</mods:dateAccessioned>
                </mods:extension>
                <mods:extension>

```

```

    <mods:dateAvailable
      encoding="iso8601">2008-11-
      02T19:12:26Z</mods:dateAvailable>
  </mods:extension>
- <mods:originInfo>
  <mods:copyrightDate
    encoding="iso8601">2007</mods:copy
    rightDate>
  </mods:originInfo>
- <mods:originInfo>
  <mods:dateIssued
    encoding="iso8601">2008-11-
    02T19:12:26Z</mods:dateIssued>
  </mods:originInfo>
- <mods:extension>
  <mods:dateSubmitted
    encoding="iso8601">2007-09-
    24</mods:dateSubmitted>
  </mods:extension>
  <mods:identifier
    type="uri">http://hdl.handle.net/10318/14
    8</mods:identifier>
  <mods:abstract>Práce pojednává o základních
    fyzikálních vlastnostech barev, principu jejich
    vnímání z hlediska fyziologie vidění,
    poukazuje na psychosomatické vlivy
    jednotlivých barev na lidskou osobnost.
    Podrobněji sleduje barvy jako nástroje
    obrazové skladby, a jejich využití v
    kinematografii, barevnou harmonii a
    barevnou kompozici. Jejím závěrem je
    psychologický, psychofyziologický a
    symbolický rozbor některých vybraných
    barev.</mods:abstract>
- <mods:physicalDescription>
  <mods:form>application/msword</mods:for
  m>
  </mods:physicalDescription>
- <mods:language>
  <mods:languageTerm
    authority="rfc3066">cze</mods:lang
    uageTerm>
  </mods:language>
- <mods:originInfo>
  <mods:publisher>Knihovna Filmové fakulty
    Akademie múzických
    umění</mods:publisher>
  </mods:originInfo>
  <mods:relatedItem
    type="constituent">http://hdl.handle.net/1
    0318/145</mods:relatedItem>
  <mods:relatedItem
    type="constituent">http://hdl.handle.net/1
    0318/149</mods:relatedItem>
  <mods:accessCondition
    type="useAndReproduction">a)
    rozmnožováním díla jakýmikoliv prostředky a
    v jakékoliv formě, včetně formy
    elektronické</mods:accessCondition>
  <mods:accessCondition
    type="useAndReproduction">b)
    bezúplatným rozšiřováním originálu Školního
    díla nebo jeho rozmnoženin pořízených podle
    písm. a)</mods:accessCondition>
  <mods:accessCondition
    type="useAndReproduction">d)
    vystavováním originálu díla nebo jeho
    rozmnoženin podle písm.
    a)</mods:accessCondition>

```

```

<mods:accessCondition
  type="useAndReproduction">e) sdělováním
  veřejnosti všemi způsoby.., zejména pak
  živým provozování díla....a zpřístupňováním v
  síti Internet...</mods:accessCondition>
<mods:subject>
  <mods:topic>Film</mods:topic>
</mods:subject>
<mods:titleInfo>Psychologické působení barev a
  jeho využití v barevné
  kinematografii</mods:titleInfo>
<mods:titleInfo
  type="alternative">Psychological Working of
  Colours and its Application in Colour
  Cinematography</mods:titleInfo>
<mods:genre>Diplomová práce</mods:genre>
</xmlData>
</mdWrap>
</dmdSec>
<amdSec ID="TMD_hdl_10318/148" />
<fileSec>
  <fileGrp USE="ORIGINAL">
    <file ID="10318_148_1"
      MIMETYPE="application/pdf" SIZE="14800"
      CHECKSUM="e5e299e104d9c4874a2183fcb0b85f6
      f" CHECKSUMTYPE="MD5"
      OWNERID="http://dspace.amu.cz/bitstream/103
      18/148/1/AMU%20test.pdf"
      GROUPID="GROUP_10318_148_1">
      <FLocat LOCTYPE="URL" xlink:type="simple"
        xlink:href="http://dspace.amu.cz/bitstream
        /10318/148/1/AMU%20test.pdf" />
    </file>
  </fileGrp>
  <fileGrp USE="TEXT">
    <file ID="10318_148_2" MIMETYPE="text/plain"
      SIZE="58"
      CHECKSUM="3ecb33c441c51c0a8f76d81e8f42024
      c" CHECKSUMTYPE="MD5"
      OWNERID="http://dspace.amu.cz/bitstream/103
      18/148/2/AMU%20test.pdf.txt"
      GROUPID="GROUP_10318_148_1">
      <FLocat LOCTYPE="URL" xlink:type="simple"
        xlink:href="http://dspace.amu.cz/bitstream
        /10318/148/2/AMU%20test.pdf.txt" />
    </file>
  </fileGrp>
</fileSec>
<structMap>
  <div />
</structMap>
</mets>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

```

Obr. 5/4: metadatový záznam diplomové práce z archivu AMU ve struktuře XML exportovaného pomocí protokolu OAI-PMH pomocí příkazu GetRecord (ve formátu mets) [převzato 2009-03-25]

Další příklad (Obr. č. 5/5) ukazuje zápis prvků standardu EVSKP-MS včetně deklarace XML s kódováním UTF-8, s úvodním prvkem evskp:metadata s definicí jmenných prostorů, tentokrát i s vnořovaným zápisem jmen fyzických osob podle standardu PersCZ a s popisem korporací podle specifikace CorpCZ. Následující příklad na obr. č. 5/5 je záznam diplomové práce z repozitáře Vysoké školy ekonomické v Praze (<http://www.vse.cz/oai>) ve formátu EVSKP-MS (oai_evskpms), PerSCZ a CorpCZ. Záznam má vnořované prvky pro fyzické

osoby a korporace (jde o jednu z možností využití formátu EVSKP-MS). V záznamu je uplatněn pouze lokální identifikátor objektu.

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd
    http://www.evskp.cz/standardy/evskp/
    http://www.evskp.cz/standardy/evskp/1.1/evskp.xsd">
  <responseDate>2009-03-27T16:10:05Z</responseDate>
  <request verb="GetRecord" metadataPrefix="oai_evskpms"
    identifier="oai:vse.cz:vskp/13101">http://www.vse.cz/oai</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:vse.cz:vskp/13101</identifier>
        <timestamp>2009-03-23T10:09:20Z</timestamp>
      </header>
      <metadata>
        <evskp:metadata version="1.1"
          xmlns:ccz="http://www.evskp.cz/standardy/corpcz/"
          xmlns:dc="http://purl.org/dc/elements/1.1/"
          xmlns:dcterms="http://purl.org/dc/terms/"
          xmlns:dctype="http://purl.org/dc/dcmitype/"
          xmlns:evskp="http://www.evskp.cz/standardy/evskp/"
          xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
          xmlns:pcz="http://www.evskp.cz/standardy/perscz/"
          xmlns:thesis="http://www.ndltd.org/standards/metadata/etdms/1.0/">
          <dc:title xml:lang="cs">Návrh systému e-shopu pro podporu
            distribuce segmentu malovýrobců českého skla.</dc:title>
          <dc:title xml:lang="en"
            evskp:typeTranslated="translated">Project of E-shop system
            for support of small producers bohemian glass segment
            distribution.</dc:title>
          <dc:creator>
            <pcz:person>
              <pcz:name>
                <pcz:foreName>Ondřej</pcz:foreName>
                <pcz:surName>Kocina</pcz:surName>
              </pcz:name>
              <pcz:dateOfBirth>1981-06-24</pcz:dateOfBirth>
            </pcz:person>
          </dc:creator>
          <dc:subject xml:lang="en">demand; simulation; e-
            shop</dc:subject>
          <dcterms:abstract xml:lang="en">Essay analyses problems of
            bohemian glass small producers segment and in the scope of
            the real firm strategy suggests their concrete solution.
            Analysis finds causes of the segment problems in
            inappropriate combination of product, market and
            distribution channels. Author finds possible solution in the
            oligopoly price leader market model. Results of the problems
            analysis are together with firm and its environment analysis
            used for SWOT analysis. Management of the firm chose S-O
            strategy, which aims sustainable development of the firm,
            differentiation of the product and enlargement and
            diversification of the market space. Results of the analysis
            are used for creation of partial strategies and plans:
            marketing mix and information strategy. Information
            strategy includes concept of the new information and
            communication system suggestion. Suggested system is
            based on hosted web e-shop application and internet
            communication software component. System enables to
            flexible react on the changes of firm strategy and plans, and
            enables to create individualized and individual services.
            Chosen architecture is ideal for the rise of virtual teams
```

```

    System should help to emerge integration and synergic
    effects.</dcterms:abstract>
<dc:publisher>
  <ccz:universityOrInstitution>
    <ccz:name xml:lang="cs">Vysoká škola ekonomická v
      Praze</ccz:name>
    <ccz:email>webmaster@vse.cz</ccz:email>

    <ccz:homepage>http://www.vse.cz/</ccz:homep
      age>
    </ccz:universityOrInstitution>
  </dc:publisher>
<dc:contributor thesis:role="advisor">
  <pcz:person>
    <pcz:name>
      <pcz:foreName>Stanislava</pcz:foreName>
      <pcz:surName>Mildeová</pcz:surName>
    </pcz:name>
    <pcz:note>Ing., CSc.</pcz:note>
  </pcz:person>
</dc:contributor>
<dc:contributor thesis:role="referee">
  <pcz:person>
    <pcz:name>
      <pcz:foreName>Jaroslav</pcz:foreName>
      <pcz:surName>Jandoš</pcz:surName>
    </pcz:name>
    <pcz:note>prof. Ing., CSc.</pcz:note>
  </pcz:person>
</dc:contributor>
<dcterms:created>2008-12-15</dcterms:created>
<dcterms:dateSubmitted>2009-03-
  23</dcterms:dateSubmitted>
<dcterms:dateAccepted>2009-01-21</dcterms:dateAccepted>
<dcterms:modified>2009-03-02T06:52:00Z</dcterms:modified>
<dc:type xml:lang="cs" evskp:typeType="TypVSKP">Diplomová
  práce</dc:type>
<dcterms:medium>application/pdf</dcterms:medium>
<dc:identifier>http://www.vse.cz/vskp/eid/13101</dc:ide
  ntifier>
<dc:language>cs</dc:language>
<thesis:degree>
  <thesis:name>Ing.</thesis:name>
  <thesis:level xml:lang="cs">Magisterský navazující
    studijní program</thesis:level>
  <thesis:discipline xml:lang="cs">Aplikovaná
    informatika/Informační
    management</thesis:discipline>
  <thesis:grantor>
    <ccz:universityOrInstitution>
      <ccz:name xml:lang="cs">Vysoká škola
        ekonomická v Praze</ccz:name>
    </ccz:universityOrInstitution>
  </thesis:grantor>
</thesis:degree>
<evskp:contact contactID="3190" />
<evskp:fileNumber>4</evskp:fileNumber>
<evskp:fileProperties fileID="108311" fileType="thesis"
  fileName="ISIS_14202_xkoco05.pdf" fileSize="8631219"
  format="application/pdf">Hlavní
  práce</evskp:fileProperties>
<evskp:fileProperties fileID="108314" fileType="appendix"
  fileName="ISIS_1201_xkoco05.pdf" fileSize="37079"
  format="application/pdf">Veřejná
  příloha</evskp:fileProperties>
<evskp:fileProperties fileID="108313"
  fileType="advisorReview" fileName="ISIS_8658_jandos.pdf"
  fileSize="46714"

```



```

        format="application/pdf">Oponentura</evskp:fileProperties
    >
    <evskp:fileProperties fileID="108312" fileType="refereeReview"
        fileName="ISIS_14202_mildeova.doc" fileSize="54784"
        format="application/msword">Hodnocení
        vedoucího</evskp:fileProperties>
    <evskp:transfer accessRights="domain"
        fileID="108311">http://www.vse.cz/vskp/id/108311</evs
        kp:transfer>
    <evskp:transfer accessRights="domain"
        fileID="108314">http://www.vse.cz/vskp/id/108314</evs
        kp:transfer>
    <evskp:transfer accessRights="domain"
        fileID="108313">http://www.vse.cz/vskp/id/108313</evs
        kp:transfer>
    <evskp:transfer accessRights="domain"
        fileID="108312">http://www.vse.cz/vskp/id/108312</evs
        kp:transfer>
    <evskp:server>
    <ccz:universityOrInstitution>
        <ccz:name xml:lang="cs">Vysoká škola ekonomická v
        Praze</ccz:name>
        <ccz:place>Praha</ccz:place>
    <ccz:department>
        <ccz:name xml:lang="cs">Centrum informačních
        a knihovnických služeb</ccz:name>
        <ccz:email>webmaster@vse.cz</ccz:email>

        <ccz:homepage>http://ciks.vse.cz/</ccz:ho
        mepage>
    </ccz:department>
    </ccz:universityOrInstitution>
    </evskp:server>
    <evskp:modified>2009-03-02T06:52:00Z</evskp:modified>
</evskp:metadata>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

```

Obr. 5/5: metadatový záznam diplomové práce z archivu AMU ve struktuře XML exportovaného pomocí protokolu OAI-PMH pomocí příkazu GetRecord (ve formátu oai_evskpms) [převzato 2009-03-25]

5.4.5 Závěr

Uvedené tři příklady dokladují platnost navržených a schválených standardů pro popis a komunikaci VŠKP a fyzických osob a korporací. Záznamy v těchto specifikacích a struktuře mohou být automaticky komunikovány přes sítě. Uvedené dva subjekty (AMU a VŠE) již poskytují metadata také do národního metadatového registru VŠKP (NR VSKP) v Brně na MU. Standardy EVSKP-MS, v. 1.1, PersCZ, v. 1.0 a CorpCZ, v. 1.0 jsou schválenými standardy i pro NR VSKP. Účastnické školy mohou však používat dle pokynů správců NR VSKP také předchozí vývojovou specifikaci EVSKP-MS, v. 0.1.

Ve uvedených národních standardech se počítá s uplatněním – v optimální míře – s nasazením trvalých identifikátorů pro VŠKP samotné a také pro fyzické osoby a korporace – po vzoru ze zahraničí. Školy, které mají nasazeny systémy, jako je DSpace (AMU aj.), využívají identifikátoru HDL amerického systému The Handle System. Většina českých systémů však zatím uplatňuje jen lokálně platné identifikátory. Čeká se na zavedení nějakého systému trvalé identifikace digitálních dokumentů na úrovni Národní knihovny ČR, který by poté mohl české vysoké školy využít. Následující části představují možné trvalé identifikátory ze světové praxe.

5.5 Vybrané trvalých identifikátory pro digitální objekty (dokumenty)

5.5.1 Úvod

Identifikace informačních entit jakéhokoli druhu či typu patří v současném období k velmi významným informačním operacím. Její charakter se v posledních 15 letech proměnil, její význam vzrůstá zejména v provozech informačních systémů nově budovaných a rozvíjených v prostředí sítě Internet. Co nejpřesnější identifikace všech potřebných informačních entit (objektů, zdrojů) je podstatnou podmínkou rychlého a efektivního provozu těchto systémů a jejich vzájemného propojování. Již dávno nepostačují k identifikaci, jež je zpravidla vždy součástí dalších informačních procesů a operací, jako jsou zjišťování, získávání, zpracování, vyhledávání a zpřístupňování nebo třeba i prodej informací, jenom **tradiční informační znaky** informačních entit (autor, název, nakladatel, rok vydání, název edice apod.). Společnost již několik desetiletí využívá k přesnější identifikaci těchto entit specifické identifikační informační znaky, jimž se jednoduše říká **identifikátory**. V podobě specificky tvořených či generovaných řetězců (skládajících se číslic, písmen nebo symbolů) **zastupují definovanou informační entitu** v různých informačních procesech. Obecně platí, že identifikátor je nutné chápat jako konvenci či jako sociální dohodu a konkrétní hodnota identifikátoru je pak závislá na charakteru organizace, která ho vytváří, udržuje a přiděluje [COYLE, 2006, s. 428]. Prakticky ve všech informačních systémech se v současnosti využívá velkého množství identifikátorů lokálního významu. V rovině mezinárodní komunikace informací v digitální době pak na značném významu nabývají takové identifikátory, které jsou v globálním rámci **jednoznačné** či **jedinečné** a zejména **trvalé** (angl. persistent). Hovoříme o **trvalých identifikátorech** informačních entit. Hrají důležitou roli v zabezpečení potřebné interoperability informačních systémů v prostředí počítačových sítí.

K již dobře známým identifikátorům mezinárodního charakteru (ISBN, ISSN) přibyla řada dalších, které pomáhají jednoznačně a trvale identifikovat nejen další typy tradičních bibliografických či knihovních entit (ISMN, ISRN, EAN), ale i novodobé **netradiční bibliografické entity**, jakými jsou digitální zdroje zpřístupňované v prostředí počítačových sítí (DOI, URN, PURL, ARK aj.), nebo také jiné informační entity dosud běžně nezpracovávané, jakými jsou **abstraktní tvůrčí díla** (ISWC, ISAN, ISTC aj.), **vyjádření tvůrčích děl** (ISRC) nebo sami **tvůrci, výrobci a distributoři informací** (ISADN, ISNI aj.).

Identifikátory můžeme z hlediska jejich struktury a obsahu rozdělit na dva základní typy:

- **Identifikátory jednoduché struktury**

Slouží jenom k jednoznačnému označení informační entity (objektu, zdroje). Tento typ identifikátoru nepodává o entitě žádnou jinou obsahovou informaci, je v podstatě jenom jednoduchým pořadovým číslem. Nicméně v určitých případech lze z čísla jisté informace vyvozovat. Tento typ identifikátoru je **trendem posledních let**. Nejnovější identifikátory z produkce ISO takového typu jsou. Tyto identifikátory bývají proto často spojovány s určitou **množinou povinných metadat**, jež jsou společně ukládána v registračních databázích (viz mezinárodní registrační systémy ISSN, ISWC, ISAN, DOI aj.).

- **Identifikátory komplexní povahy**

Jsou někdy nesprávně označovány jako „inteligentní“. Slouží nejenom k jednoznačné identifikaci informační entity, ale navíc o ní podávají i určité významové informace (například číslo ISBN identifikuje zemi, jazyk nebo geografickou oblast, v níž je kniha vydána, nebo nakladatele, jemuž byl přidělen kód).

K trvalým identifikátorům pro digitální objekty patří:

DOI	Digital Object Identifier
ARK	Archival Resource Key
HDL	Handle
PURL	Persistent URL (Uniform Resource Locator)
URN	Uniform Resource Name
URI	Uniform Ressource Identifier

5.5.2 URI – obecný rámec trvalých identifikátorů

Obecný identifikátor URI internetové veřejnosti poprvé představil Tim Berners-Lee v roce 1994 - tehdy ještě pod pojmenováním **Univerzální identifikátor zdroje** (Universal Resource Identifier), a to v rámci informativního dokumentu RFC 1630 [RFC 1630, 1994].

URI byl tehdy užíván v diskusi kolem implementace WWW [HILSE, 2006, s. 4]. Nové pojetí identifikátoru URI se změněným pojmenováním **Jednotný identifikátor zdroje** (Uniform Resource Identifier) bylo publikováno již jako internetový standard poprvé v roce 1998 [RFC 2396, 1998].

Aktuálně platná verze tohoto internetového standardu pochází z roku 2005. Je věnován obecné syntaxi identifikátoru URI [RFC 3986, 2005].

Jednotný identifikátor zdroje URI je ve standardu definován jako **řetězec znaků, který slouží k identifikaci (nebo lokalizaci, vyhledání, směrování, odkazování apod.) nějakého [informačního] zdroje. Může jím být zdroj abstraktní nebo fyzický.** Zdrojem (entitou, objektem) může být podle standardu prakticky cokoliv: webová stránka, dokument, databáze, obraz, služba, kolekce (fond) zdrojů, fyzický objekt (fyzická osoba, automobil aj.), událost, pojem aj. [URI Resource pages, 2007]. Zdrojem může být i metadatový záznam. **URI je obecným rámcem identifikátorů informačních zdrojů.**

Jednotlivé typy obecného identifikátoru URI se rozlišují podle svého **schématu**. Obecné schéma URI je následující:

Jméno schématu:Specifické schéma

Jméno schématu se od specifického schématu odděluje dvojtečkou (bez mezer okolo ní). Výklad specifického schématu je závislý na schématu.

Příklady identifikátorů URI:

http://http://www.ifla.org/VII/s13/garr/garr.pdf

telnet://dec59.ruk.cuni.cz

ftp://ftp.rfc-editor.org/in-notes/rfc4452.txt

mailto:Eliska.Novakova@ff.cuni.cz

info:lccn/n78089035

info:doi/10.1000.10/123456

z39.50s://melvyl.ucop.edu/cat

urn:nbn:de:gbv:089-3321552

K frekventovaným schémátům URI v současnosti patří schéma „**http**“, které hraje důležitou roli při určování lokalizace zdrojů na webu, a schéma „**info**“, které je schématem novým, zavedeným zejména pro potřeby jednoznačné identifikace některých tradičních lokálně platných identifikátorů (například známého identifikátoru záznamu katalogu Kongresové knihovny LCCN) s cílem jejich potřebného zapojení do automatického propojování informačních zdrojů v síťovém prostředí (pomocí standardu OpenURL a technologie SFX).

Ke známým schématům URI patří také schéma „urn“, tedy **Jednotné jméno zdroje** (Uniform Resource Name), které souvisí s definovaným a schváleným jmenným prostorem. Je nutné poznamenat, že na rozdíl od „urn“, není za schéma považován **Jednotný lokátor zdroje „url“** (aby bylo URL nalezené v rámci Internetu akční, je zapotřebí schématu URI „http“).

Ve výše uvedených příkladech jsou některá schémata **specifickými protokoly** („http“, „telnet“, „ftp“). Jiná schémata protokoly nejsou, ale s určitým protokolem souvisejí („z39.50s“, „mailto“). Schéma „urn“ souvisí s více protokoly. Schéma „info“, naopak, nesouvisí s protokolem žádným. Schéma „info“ slouží **pouze k identifikaci zdroje** (je čistým identifikátorem).

Všechna **schválená schémata URI** jsou v současné době v globálním rámci registrována na serveru organizace IANA:

<http://www.iana.org/assignments/uri-schemes.html>

Na serveru OCLC (<http://info-uri.info/>) jsou v globálním rámci registrována již také všechna schválená (celkem 25 v prosinci 2007) **schémata URI typu „info“**. Patří k nim (výběr):

info:ark/	(jmenný prostor pro identifikátory ARK)
info:arxiv/	(jmenný prostor pro identifikátory archivu ArXiv.org)
info:doi/	(jmenný prostor pro identifikátory DOI)
info:fedora/	(jmenný prostor pro „Fedora Digital Objects and Disseminations“)
info:hdl	(jmenný prostor pro identifikátory Handle)
info:lc/	(identifikátory Kongresové knihovny)
info:lccn/	(jmenný prostor pro identifikátory katalogizačních záznamů LC)
info:nla/	(jmenný prostor pro identifikátory Národní knihovny Austrálie)
info:oclcnum/	(jmenný prostor pro identifikátory záznamů katalogu OCLC)
info:ofi/	(jmenný prostor pro identifikátory registru pro rámec OpenURL)
info:pmid	(jmenný prostor pro identifikátory systému PubMed)
info:rlgid	(jmenný prostor pro identifikátory záznamů sítě RLG)
info:sici/	(jmenný prostor pro identifikátory SICI)
info:sid/	(jmenný prostor pro zdrojové identifikátory pro rámec OpenURL)
info:srw/	(jmenný prostor služby SRW – Search/Retrieve Web)

Schéma URI typu „info“ se skládá z (viz příklady výše v textu):

- Jména schématu „info:“
- Jmenného prostoru schématu „info“ (například „lccn“, „doi“ aj. - viz registr OCLC (<http://info-uri.info/>))
- Symbolu lomítka („/“)
- Specifického jmenného prostoru (řetězce)

Schéma „info“ vzniklo uvnitř knihovnické a vydavatelské komunity a souvisí s rozvojem výše zmíněného amerického standardu **OpenURL**, který je jedním z podstatných nástrojů efektivního propojování různorodých informačních zdrojů.

Výklad schématu URI typu „info“ je představen v internetovém standardu z roku 2006 [RFC 4452, 2006].

5.5.3 URN

5.5.3.1 URN obecně

Vznik **Jednotného jména zdroje URN** (Uniform Resource Name) je spojen také s rokem 1994, kdy byl zveřejněn internetový standard RFC 1737 s názvem „**Funkční požadavky na Jednotné jméno zdroje**“. Byly v něm specifikovány jen požadavky na tento trvalý identifikátor (bez specifikace syntaxe) [RFC 1737, 1994].

Autoři tohoto standardu stanovili (v kap. 2) následující požadavky na trvalý identifikátor URN:

- Globální záběr (dané URN v sobě nezahrnuje informaci o umístění zdroje a má všude na světě stejný význam)
- Globální jedinečnost (stejně URN nemůže být nikdy přiděleno dvěma různým zdrojům)
- Trvalost (URN je stálým identifikátorem, je platné navždy bez ohledu na existenci samotného zdroje, který identifikuje)
- Škálovatelnost (URN může být přiděleno libovolnému zdroji, který se může v síti vyskytnout - nyní nebo kdykoliv v budoucnosti)
- Podpora minulých schémat (URN musí umožnit integraci jiných již existujících jmenných schémat, pokud budou vyhovovat jiným požadavkům ve standardu uvedených; do URN mohou být integrovány například známé identifikátory publikací, jako jsou ISBN, ISSN aj.)
- Rozšiřitelnost (jakékoli schéma URN musí dovolit jeho případné budoucí rozšíření)
- Nezávislost (určování pravidel pro přidělování identifikátorů URN je zcela v kompetenci ustanovené autorizované organizace)
- Směrování (využívání URN lze realizovat pomocí síťově dostupné směrovací služby, která bude mapovat URN až k odpovídajícímu zdroji); jestliže identifikátory URN mají odpovídající URL, musí existovat mechanismus, který je bude směřovat k překladu daného URN na URL

Stejně či podobné požadavky jsou dnes kladeny ale i na jiné trvalé identifikátory.

URN je **trvalým identifikátorem informačního zdroje, který je nezávislý na jeho lokaci nebo metodě přístupu**.

Syntax identifikátoru URN je dána internetovým standardem z roku 1997 [RFC 2141, 1997]. Vychází ze syntaxe jednotného identifikátoru zdroje **URI**. (URN je definovaným a schváleným schématem URI - viz předchozí podkapitola). Syntax URN může podle uvedeného standardu zahrnovat několika hierarchicky uspořádaných jmenných prostorů (Namespaces). Obecnou strukturu definuje výše uvedený standard (kap. 2, s. 1) následujícím způsobem:

urn:NID:NSS

Schéma „urn:“ (včetně dvojtečky za ním) je povinnou částí celého identifikátoru a může být zapisováno malými nebo velkými písmeny. Zkratka **NID** označuje globálně definovaný a schválený **identifikátor jmenného prostoru** (Namespace Identifier). NID se může skládat z písmen, číslic a pomlček. Zkratka **NSS** označuje **specifický řetězec jmenného prostoru** (Namespace Specific String). Dvojtečková syntax, oddělovající identifikátor jmenného prostoru

a specifický řetězec jmenného prostoru, je podle standardu RFC 2141 povinná. NSS může mít svoji vlastní definici, která je dána jmenným prostorem (NID). NSS může zahrnovat další syntaktické dvojtečky (popř. také pomlčku), jež oddělují identifikátory dílčích jmenných prostorů (ve zkratce SNID, SubNamespace Identifier). Identifikátor NSS může zahrnovat libovolné znaky. Informativní dokument RFC 3188 pro schválenou doménu URN:NBN (určenou pro systémy národních bibliografií) byl zveřejněn v roce 2001 [RFC 3188, 2001].

Doporučuje jako součást specifického řetězce **kód země** (optimálně dvouznakový kód podle ISO 3166) a kód organizace, která informačnímu zdroji přiděluje trvalý identifikátor. Například v reálném trvalém identifikátoru URN:

urn:nbn:de:gbv:7-isbn-90-6984-508-3-8

kód „de“ reprezentuje Německo a kód „gbv:7“ instituci (Univerzitní knihovna Göttingenské univerzity), která přidělila zdroji řetězec zahrnující v tomto příkladě ISBN vydané publikace [citace přímo: HILSE, 2006]. Nicméně, z pohledu celkové syntaxe identifikátoru URN je celek specifického řetězce chápán jako jednoduchý, neprůhledný řetězec (bez významu) [URI Resource pages, 2007]. ISBN zahrnuté ve výše uvedeném identifikátoru URN:NBN by se mohlo vyskytnout také jako čistý NSS v rámci již schváleného jmenného prostoru ISBN:

urn:isbn:90-6984-508-3-8

K dnešnímu dni byla již definována a schválena řada jmenných prostorů URN. Jejich registraci zajišťuje organizace IANA (<http://www.iana.org/assignments/urn-namespaces>). Nejrozšířenější a v praxi nejužívanější je výše zmíněný jmenný prostor (doména) **URN:NBN** (je dán výše citovaným standardem RFC 3188). Schválen a registrován je také jmenný prostor **URN:ISBN** (RFC 3187). K dalším schváleným jmenným prostorům patří (ve výběru):

URN:IETF

(je definován RFC 2648 a je určen pro samotné internetové standardy RFC, které zajišťuje internetová inženýrská skupina IETF)

URN:ISAN

(definice standardu RFC se připravuje, jmenný prostor je určen pro standardní identifikátor ISAN)

URN:ISSN

(je definován RFC 3044 a je určen pro standardní identifikátor ISSN)

URN:MPEG

(je definován RFC 3614 a je určen pro zdroje, jež jsou součástí publikovaných standardů skupiny MPEG (Motion Picture Experts Group))

URN:UUID

(je definován RFC 4122 a je určen pro identifikátory UUID (Universally Unique IDentifiers))

URN:XMLORG

(je definován RFC 3120 a je určen pro zdroje pocházející z repozitáře XML:ORG organizace OASIS (Organization for the Advancement of Structured Information Standards))

Příprava definice jmenných prostorů se řídí standardem RFC 2611 z roku 1999 [RFC 2611, 1999].

Využívání tradičních identifikátorů (ISBN, ISSN aj.) v rámci URN je dáno informativním dokumentem RFC 2288 z roku 1988 [RFC 2288, 1988].

Zatímco zavedení systému na generování a přidělování trvalých identifikátorů URN není složitou záležitostí (takový generátor je například k dispozici také v rámci českého WebArchivu v NK ČR), vybudování vlastního systému pro jejich účinné **směrování** (angl. „resolution“) v prostředí současné sítě Internet problémem být může, určitě pak finančním. Vybudování vlastní síťově dostupné **směrovací služby** na národní, regionální, lokální, oborové nebo jiné úrovni (angl. „resolution service“) je nákladnou záležitostí. Směrovací služba na globální či mezinárodní úrovni zatím vybudována není.

Z hlediska koncepčního znamená **směrování identifikátoru URN jeho mapování k jedné nebo více adresám URL**, na kterých se informační zdroj nalézá. Když uživatel aktivuje nějaké URN (kliká na něj), internetový prohlížeč najde na serveru směrovací služby v příslušném záznamu seznam odpovídajících adres URL, vybere jednu adresu a pokusí se zdroj vyhledat. Pokud se první pokus nezdaří, prohlížeč se pokusí zdroj vyhledat na základě další adresy URL. Celý proces zůstává v těchto krocích pro uživatele transparentní. Jestliže se informační zdroj objeví v kopii na dalším serveru, je do záznamu na serveru směrovací služby přidána nová adresa URL. Jestliže je informační zdroj z nějakého serveru vymazán, musí být příslušná adresa URL ze záznamu směrovací služby vymazána také. Jestliže je zdroj přesunut na jiné místo, URL v záznamu směrovací služby musí být aktualizováno. Ve všech případech ale **URN zůstává vždy stejné** [URI Resource pages, 2007].

K dosažení co nejjednodušší implementace směrovací služby URN se v praxi využívá jednoduchého mechanismu směrování z URN k URL, z URN ke zdroji samotnému a popřípadě také z URL k odpovídajícímu URN, existuje-li. Jde o **Triviální směrovací protokol HTTP** (THTTP, Trivial HTTP protocol), který využívá protokol HTTP jako základu pro síťový přístup ke směrovací službě. Je definován standardem RFC 2169. Protokol specifikuje syntax dotazu a odpovědi. Ve formulaci dotazu je zahrnuto textové označení směrovací služby (serveru), typ směrovacího dotazu a vlastní identifikátor, který má být směrován. Odpověď je závislá na typu dotazu a může zahrnovat jedno nebo více URL nebo URN, popřípadě jeden nebo více informačních zdrojů nebo metadat, která zdroje popisují [HILSE, 2006, s. 12-13].

5.5.3.2 Jmenný prostor URN:NBN

NBN je schváleným jmenným prostorem URN pro potřeby **systémů národních bibliografií**. Byl připraven a zaregistrován v roce 2001 Národní knihovnou Finska (v rámci informativního dokumentu RFC 3188). Autorem textu je pracovník této knihovny Juha Hakala. Identifikátory typu URN:NBN začaly být přidělovány digitálním zdrojům postupně již od roku 1998 v národních knihovnách Finska, Švédska, Norska, Německa aj. v souvislosti se získáváním, ukládáním a zpřístupňováním elektronických dokumentů v rámci systémů národních bibliografií. Globální registraci jmenných prostorů přidělených v rámci domény URN:NBN zajišťuje Kongresová knihovna ve Washingtonu (Library of Congress, <http://www.loc.gov>). Zainteresované národní knihovny jsou odpovědné za přidělování a registraci identifikátoru dílčího jmenného prostoru, kterým je v optimálním případě dvouznakový kód země specifikovaný v mezinárodní normě ISO 3166. Za kódem země se může a nemusí vyskytovat identifikátor dalšího dílčího jmenného prostoru, reprezentujícího instituci či organizaci přidělující kódy NSS. Registraci dalších dílčích jmenných prostorů zajišťují národní knihovny. Registry těchto domén musejí být dostupné na WWW. Kódy NSS by měly být od předchozího kódu podle informativního dokumentu RFC 3188 odděleny pomlčkou.

Příklady:

urn:nbn:fi-fe19981001

urn:nbn:de:gbv:7-isbn-90-6984-508-3-8

Některé národní knihovny již také připravily **systémy směrovacích služeb** pro svoji doménu URN:NBN. Značné pokročilosti dosáhl systém směrovací služby, který buduje Národní německá knihovna ve Frankfurtu nad Mohanem v rámci **projektu EPICUR**. Jeho webová prezentace je k dispozici na WWW (<http://www.persistent-identifier.de>) [KORB, 2006]. Hlavním typem elektronických dokumentů, kterým je v současnosti identifikátor URN v Německu přidělován, jsou **disertační práce**. Systém elektronických disertací je budován v rámci jiného národního projektu zvaného **DissOnline** (<http://www.dissonline.de>).

Německá národní knihovna zahájila v září 2001 proces decentralizovaného přidělování URN elektronickým disertacím, které dostává v rámci povinného výtisku (registrace disertací je nedílnou součástí systému souběžné německé národní bibliografie). Vedle disertací jsou však identifikátory URN přidělovány již také jiným typům elektronických dokumentů, zatím dokumentům **statické povahy**, jako jsou **knižní publikace** komerčních nakladatelů nebo jiných vydavatelských subjektů, může jít i o mezinárodní organizace (viz například volně dostupná publikace německých autorů, publikovaná konsorciem CERL, která je citovaná v tomto výukovém textu [cit. přímo: HILSE, 2006]). Řešení pro přidělování URN dynamickým elektronickým zdrojům (například online seriálovým publikacím a jejich článkům) se připravuje [KORB, 2006].

The screenshot shows the Deutsche Nationalbibliothek search interface. The search results are displayed in a table format. The first result is a dissertation titled "Combination methods for automatic document organization [Elektronische Ressource] / Stefan Siersdorfer". The metadata for this entry includes the author (Siersdorfer, Stefan), the year (2006), the format (Online-Ressource), the university (Saarbrücken, Univ.), the type (Diss.), the year (2005), the persistent identifier (urn:nbn:de:bsz:291-scidok-4956), the URL (Archivserver der Deutschen Nationalbibliothek), the DDC notation (006.31 [DDC22ger] ; 025.344 [DDC22ger]), and the subject group (004 Informatik ; 020 Bibliotheks- und Informationswissenschaft).

Download	DDC-Suche	Wechsel zum Archivalienkatalog des Deutschen Exilarchivs 1933-45	Ende
<p>Ihre Aktion gefiltert suchen [und] (Sachgruppe [sgt]) 020 Filtereinstellungen 18</p> <p>eingrenzen (Hochschulschrift [hss]) diss Filtereinstellungen</p>			
<p>Titel: Combination methods for automatic document organization [Elektronische Ressource] / Stefan Siersdorfer</p> <p>Verfasser: Siersdorfer, Stefan</p> <p>Erscheinungsjahr: [2006]</p> <p>Umfang/Format: Online-Ressource</p> <p>Hochschulschrift: Saarbrücken, Univ., Diss., 2005</p> <p>Persistent Identifier: urn:nbn:de:bsz:291-scidok-4956</p> <p>URL: Archivserver der Deutschen Nationalbibliothek</p> <p>DDC-Notation: 006.31 [DDC22ger] ; 025.344 [DDC22ger]</p> <p>Sachgruppe: 004 Informatik ; 020 Bibliotheks- und Informationswissenschaft</p>			

Obr. č. 5/: Záznam disertace v katalogu Německé národní knihovny ve výstupním formátu systému PICARTA s trvalým identifikátorem URN [2007-11-30]

Celý systém URN:NBN v Německu z hlediska organizačního a technického zajišťuje Německá národní knihovna. Systém zahrnuje kompletně jak přidělování, tak směrování identifikátorů URN. Identifikátory URN jsou přidělovány dokumentům, které jsou již uloženy

v digitálním repozitáři Německé národní knihovny (NNK) nebo v něm budou uloženy v budoucnu (zejména online dokumenty). Přidělovány jsou také dokumentům, které jsou ukládány v důvěryhodných (certifikovaných) repozitářích jiných institucí, jež jsou v rámci domény URN:NBN zaregistrovány. Identifikátory URN jsou dokumentům přidělovány bez ohledu na to, jestli jim již byl přidělen dříve jiný standardní identifikátor (ISBN, ISSN aj.) nebo jiný trvalý identifikátor (DOI, PURL aj.). Na bázi smluvní mezi NNK a dalšími subjekty je nezbytné dohodnout neustálou aktualizaci adres URL ukládaných v centrální databázi směrovací služby.

Celý systém URN:NBN, v jehož rámci bylo od září 2001 do konce roku 2007 přiděleno již téměř **117 000 URN**, zahrnuje několik hierarchicky vystavěných komponent. Jmenný prostor URN:NBN zahrnuje v současnosti již více geograficky podřízených jmenných prostorů z oblasti německy hovořících zemí („de“ pro SRN, „ch“ pro Švýcarsko a „at“ pro Rakousko). Dokumentace systému deklaruje otevřenost i pro další zájemce z jiných zemí.

V celém systému URN:NBN je v současné době již registrováno 287 institucí, 273 z nich se aktivně účastní přidělování URN. Identifikátory URN jsou přidělovány v lokálních systémech, Německá národní knihovna pak zajišťuje jejich centrální registraci (v rámci metadatových záznamů) a jejich směrování. Je zodpovědná a spravuje také přidělování identifikátorů dílčích jmenných prostorů. Následují příklady vybraných identifikátorů dílčích jmenných prostorů:

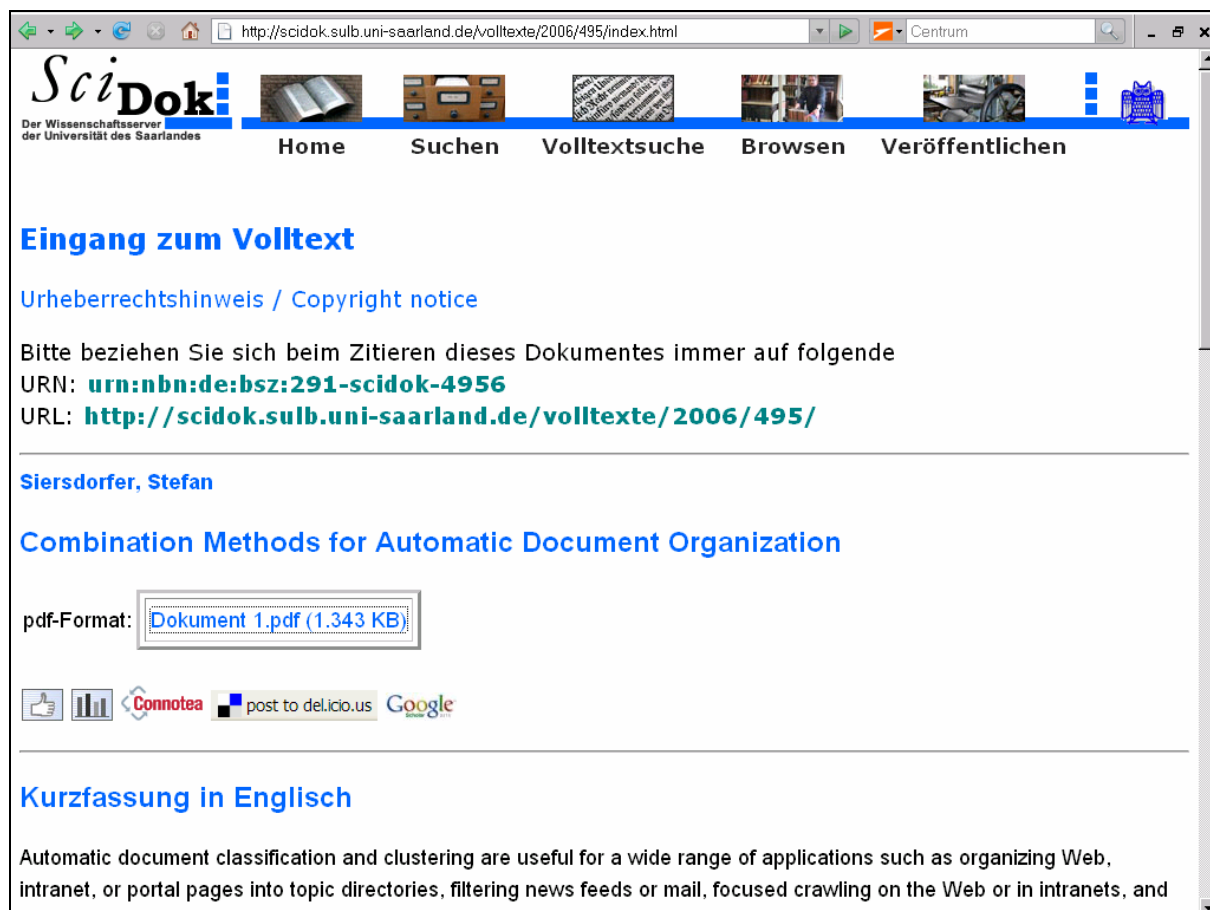
Deutsche Nationalbibliothek	urn:nbn:de:101
Bayerische Staatsbibliothek	urn:nbn:de:bvb:12
Universitätsbibliothek der Humboldt-Universität zu Berlin	urn:nbn:de:kobv:11
Saarländische Universitäts- und Landesbibliothek	urn:nbn:de:bsz:291
Niedersächsische Staats- und Universitätsbibliothek Göttingen	urn:nbn:de:gbv:7
Universitätsbibliothek Bielefeld	urn:nbn:de:0070
Universitätsbibliothek Regensburg	urn:nbn:de:bvb:355
Schweizerische Landesbibliothek	urn:nbn:ch:slb
Schweizerische Landesbibliothek	urn:nbn:ch:bel
Österreichische Nationalbibliothek	urn:nbn:at:123
Story Verlag Berlin	urn:nbn:de:0058
Universitätsverlag Karlsruhe	urn:nbn:de:0072
EXLIBRIS	urn:nbn:de:aleph
Gutenberg-Museum Mainz	urn:nbn:de:0129

Německé knihovny, které se účastní systému URN:NBN, získávají identifikátor dílčího jmenného prostoru podle svého členství v regionální knihovní síti. Například identifikátor „urn:nbn:de:gbv:089-3321752945“ patří dokumentu, který je uložen v repozitáři Hannoverské univerzity, jejíž Univerzitní knihovna (kód 089) patří do Společné knihovní sítě (Gemeinsamer Bibliotheksverbund, kód gbv).

Disertační práce, prezentovaná v katalogizačním záznamu Německé národní knihovny (viz obr. č. 5/6), získala trvalý identifikátor „urn:nbn:de:bsz:291-scidok-4956“. Úplný text je primárně uložen v univerzitním repozitáři zvaném SciDok, který je provozován na Sárské univerzitě, jejíž Univerzitní knihovna (kód 291) patří do působnosti Knihovnické služby – Centra Baden-Württemberg (Bibliotheksservice-Zentrum Baden-Württemberg, kód bsz).

Na obrázku č. 5/7 je vidět metadatový záznam stejné disertace jako na obr. č. 5/6. Jde o záznam z repozitáře „SciDok“ Sárské univerzity, na který bude uživatel vždy automaticky směřován z výchozího URN (nechť se URN vyskytuje kdekoliv, v katalogu NNK, dokumentačním serveru univerzity aj.). Reálná URL adresa metadatového záznamu této elektronické disertace na serveru SciDok je:

<http://scidok.sulb.uni-saarland.de/volltexte/2006/495/index.html>



Obr. č. 5/7: Výstupní metadatový záznam disertační práce z certifikovaného repozitáře SciDok Sárské university s trvalým identifikátorem URN i URL [2007-11-30]

Úplný text této disertace lze získat (pomocí odkazu [Dokument 1.pdf (1.343 KB)]) primárně ze stejného serveru SciDok. Jeho URL adresa je modifikací URL adresy metadatového záznamu:

http://scidok.sulb.uni-saarland.de/volltexte/2006/495/pdf/Dissertation_901_Sier_Stef_2005.pdf

Ze záznamu disertace autora Stefana Siersdorfera z katalogu Německé národní knihovny (viz obr. č. 5/6) může uživatel také volit další lokalizaci (URL) stejné disertace (digitální kopie), a sice URL na depozitním serveru NNK (povinný výtisk). Kliknutím získá nejdříve zkrácený metadatový záznam (<http://deposit.d-nb.de/cgi-bin/dokserv?idn=978586093>), ze kterého lze postoupit k úplnému záznamu (ve formátu METADISS) nebo k úplnému textu ve formátu PDF. Pod hyperlinkem MD5-Fingerprint uživatel najde digitální signaturu elektronické disertace, která chrání jeho integritu:

fecb49c0d639fac5dda08200ebc31843 978586093.pdf

Identifikátory URN, které jsou elektronickým dokumentům přidělovány v Německu v rámci domény URN:NBN, musejí být včetně příslušných URL registrovány na centrálním serveru směrovací služby do **24 hodin od doby jejich zveřejnění na lokálním dokumentačním serveru.**

NNK v rámci **centrální směrovací služby zajišťuje řadu funkcí:** rozhraní pro registraci URN a odpovídajících URL, rozhraní pro aktualizaci URL adres, pravidelnou automatickou kontrolu platnosti všech uložených URL adres, pravidelnou automatickou kontrolu integrity či autenticity uložených digitálních dokumentů prostřednictvím digitální signatury MD5-fingerprintu, zveřejnění a aktualizaci seznamu registrovaných identifikátorů dílčích jmenných prostorů na serveru směrovací služby, konfigurovatelné směrování identifikátorů URN a údržbu stabilní technické infrastruktury celého systému.

Jednotlivé instituce a organizace registrované v systému URN:NBN zajišťují: nastavení standardního směrování URN, informace o tom, zda-li je URN spojeno s jednou nebo více adresami URL, přidělování URN jednotlivým dokumentům (jako celkům nebo jejich částem, jejich verzím), kontrolu autenticity uložených digitálních dokumentů.

5.5.4 HDL (Handle)

Zkratkou HDL bývá v systémech digitálních archivů (viz např. obr. č. 5/9 dále v této podkapitole) označován jeden z nejznámějších trvalých identifikátorů - identifikátor **Handle** (anglické slovo „Handle“ se do češtiny nepřekládá). Identifikátor Handle je americkou variantou identifikátoru URN (jeho komplexní systémové zabezpečení je však jeho předností a výhodou). Handle byl vyvinut v rámci uceleného systému **The Handle System** (<http://www.handle.net>), který založila a spravuje americká **Korporace pro národní výzkumné iniciativy CNRI** (Corporation for National Research Initiatives, <http://www.cnri.reston.va.us/>). Identifikátor Handle byl původně připraven pro projekt digitálního archivu pro technické zprávy z oboru počítačové CSTR (Computer Science Technical Reports). Jeho vývoj tehdy financovala agentura DARPA (Defense Advanced Research Projects Agency). Systém Handle byl navržen výlučně jako jmenná komponenta systému pro zajištění přístupu k digitálním objektům, které byly uloženy v digitálních repozitářích. První implementace systému Handle byla k dispozici na podzim 1994 [HILSE, 2006, s. 17].

Celý Systém Handle zahrnuje různé prvky: **soubor protokolů, definovaný jmenný prostor a implementaci referenčního softwaru.** Handle je identifikátorem, který je vytvářen Systémem Handle, jenž vyhovuje definici jmenného prostoru Systému Handle. Soubor protokolů definuje také protokol, který zajišťuje směrování instituce či organizace, která identifikátor Handle přidělila, a také výměnu autentikačních informací určených pro různé úlohy s ohledem na správu dat připojených k identifikátoru Handle. Protokol definuje mechanismus, který umožňuje lokalizovat instituci či organizaci, která má na starosti informace vztahující se k pojmenovanému zdroji.

K hlavním cílům Systému Handle patří:

- Jedinečnost identifikátorů Handle
- Trvalost (stálé a operativní propojení mezi identifikátorem Handle a identifikovanou entitou je udržováno uvnitř Systému Handle; trvalost je tedy v tomto systému funkcí administrativního charakteru a péče)
- Vícenásobné výskyty (HDL je schopný odkazovat na vícenásobné výskyty pojmenovaného informačního zdroje)

- Rozšířitelný jmenný prostor (instituce, která přiděluje identifikátory HDL, si může do Systému Handle zavést svůj vlastní jmenný prostor - integrace v podobě dílčího jmenného prostoru je možná)
- Mezinárodní podpora (HDL je založen na kódování Unicode 2.0)
- Model distribuované služby (jedna globální služba může postoupit dotazy na Handle do lokální služby zainteresované instituce, ale může dotaz zodpovědět i sama)
- Zabezpečená jmenná služba (operace v rámci databáze Handle musejí být autorizovány)
- Distribuovaná administrativní služba (autorizační mechanismus se využívá individuálně pro každý identifikátor Handle, to znamená, že administrace může být distribuována individuálně)
- Efektivní směrovací služba (administrativní protokol je oddělen od protokolu směrovacího, což umožňuje snadnější distribuci požadovaných zdrojů)

Jmenné schéma (struktura) identifikátoru Handle zahrnuje dvě hlavní části, které jsou odděleny syntaktickým lomítkem (/).

Příklady:

cnri-1/1995.02.12.16.42.21;9

1012.Theses/1998-Charland.Nathalie(1970)-[HTML]

1721.1/14399

Část jména před lomítkem (prefix) identifikuje jednoznačně **instituci přidávající identifikátory** (NA, Handle Naming Authority). Jméno NA je přidělováno Globální službou Handle. V současné době jsou jména NA vyjadřována jen pomocí **desetinné číselné notace** (viz reálné jméno 1721.1 v posledním příkladu výše; vyskytuje se na obr. č. 5/9). Znak tečky (.) je využíván k vyjádření cesty v hierarchii jmen NA (cesta se čte zleva doprava). Z historických důvodů se však vyskytují i jména složená z písmen a číslic (viz první a druhý výše uvedený příklad, prezentující dříve přidělované identifikátory).

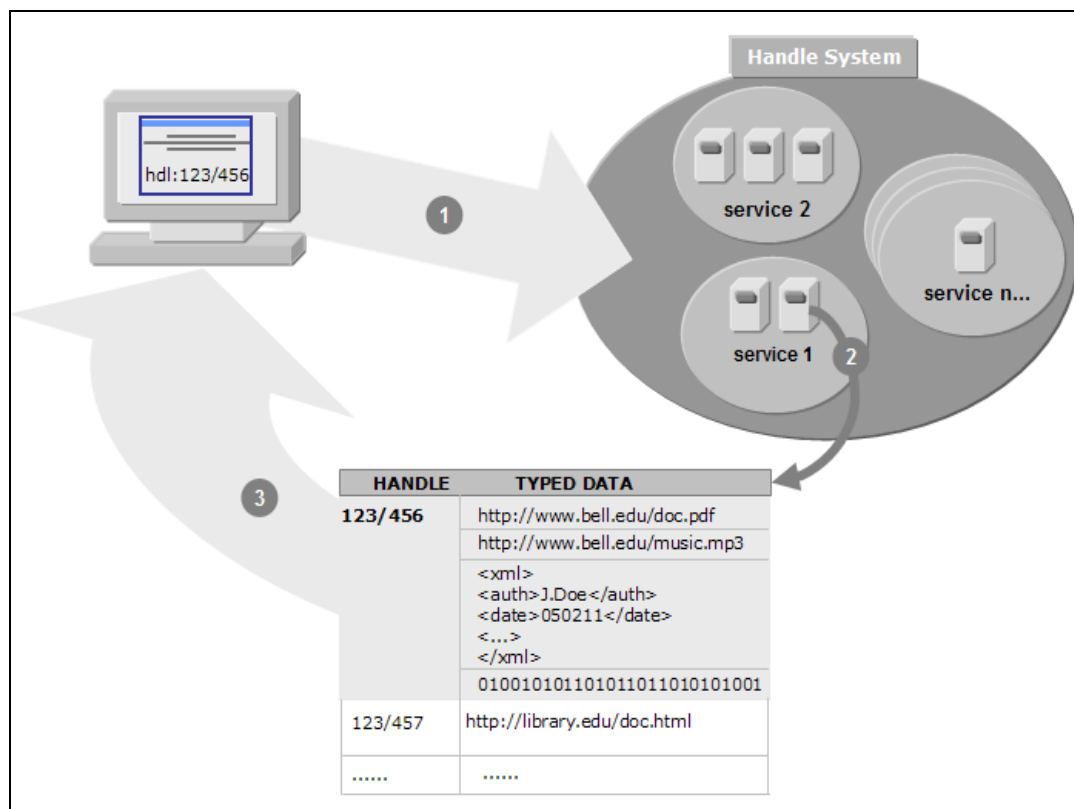
Část jména za lomítkem (sufix) identifikuje **jméno konkrétního digitálního objektu** (Handle Local Name). Lokální jméno přiděluje daná instituce (NA). V podstatě neexistují žádné limity ke strukturování lokálního jména. Jméno ale musí být vyjádřeno prostřednictvím znaků definovaných Univerzálním souborem znaků UCS-2 (Unicode, verze 2.0).

Systém Handle je navržen ke směrování identifikátorů HDL (schéma směrování - viz obr. č. 5/8) nezávisle na Systému doménových jmen (DNS). Ke směrování je využíván vlastní server **Globálního systému Handle**, který zajišťuje CNRI. Globální server spravuje **databázi institucí či korporací** (Naming Authorities), které přidělují identifikátory v rámci své působnosti. Každá instituce může vytvořit svoji vlastní infrastrukturu. Systém Handle umožňuje vytváření zrcadel a delegování směrovacích služeb na jiné servery Handle.

Systém Handle také nemusí spoléhat na adresy URL. Může jich ale využívat prostřednictvím proxy serveru, který umí přeměňovat adresy URL na základě uložení těchto adres v rámci záznamu identifikátoru HDL.

V rámci lokálních serverů jsou digitálním zdrojům přidělována konkrétní HDL a zároveň je zajišťována a neustále aktualizována vazba mezi nimi a příslušnými URL. Program k zajištění uvedených procesů lze stáhnout ze sídla systému „The Handle System“ (Local Handle System Server-JAVA Version). Součástí celého programového balíku je i plug-in (Handle System

Resolver Plug-In), který zajišťuje směrovací službu (k cílovému vyhledání samotného objektu).



Obr. č. 5: Schéma směrovací služby systému The Handle System na základě identifikátoru HDL [převzato z WWW dokumentace, 2007-11-30]

Americký systém „The Handle System“ (<http://www.handle.net/>) je v současné době využíván různými institucemi či korporacemi a je aplikován v řadě projektů a systémů. Stal se základem významného mezinárodního identifikačního systému DOI (<http://www.doi.org>), který je využíván obrovským počtem komerčních vydavatelů. (Proxy server pro směrování DOI identifikátorů je k dispozici veřejně na URL: <http://dx.doi.org/>). Systém Handle je integrován také ve známém volně dostupném softwaru **DSpace**, který v současné době využívá velké množství institucionálních repozitářů po celém světě (viz příklad identifikátoru digitálního objektu na obr. č. 5/9).

Systém Handle reprezentuje **spolehlivý identifikační mechanismus** a není pravděpodobné, že by v blízké budoucnosti přestal existovat [HILSE, 2006, s. 19]. Zájemcům o Systém Handle CNRI nabízí převzetí potřebného softwaru zdarma. Od června 2006 je však zpoplatněn registrační proces (50 USD) a za roční služby se platí stejná částka (50 USD).

Na obr. č. 5/9 je příklad metadatového záznamu disertační práce uložené v institucionálním repozitáři spravovaném Knihovnamí Massachusettského technologického institutu (MIT). Repozitář je provozován v softwaru DSpace (<http://dspace.mit.edu>). Metadatový záznam a úplný text elektronické disertace jsou identifikovány jako celek (kontejnerové řešení uložení celého digitálního objektu) jedinečným trvalým identifikátorem Handle (1721.1/14399). Prostředí digitálního repozitáře vybízí uživatele k jednoznačnému citování prostřednictvím zápisu URL adresy, která zahrnuje identifikátor HDL i jméno směrovacího serveru:

<http://hdl.handle.net/1721.1/14399>

Úplný text se nachází na reálné URL adrese:

<http://dspace.mit.edu/bitstream/1721.1/14399/1/20143315.pdf>



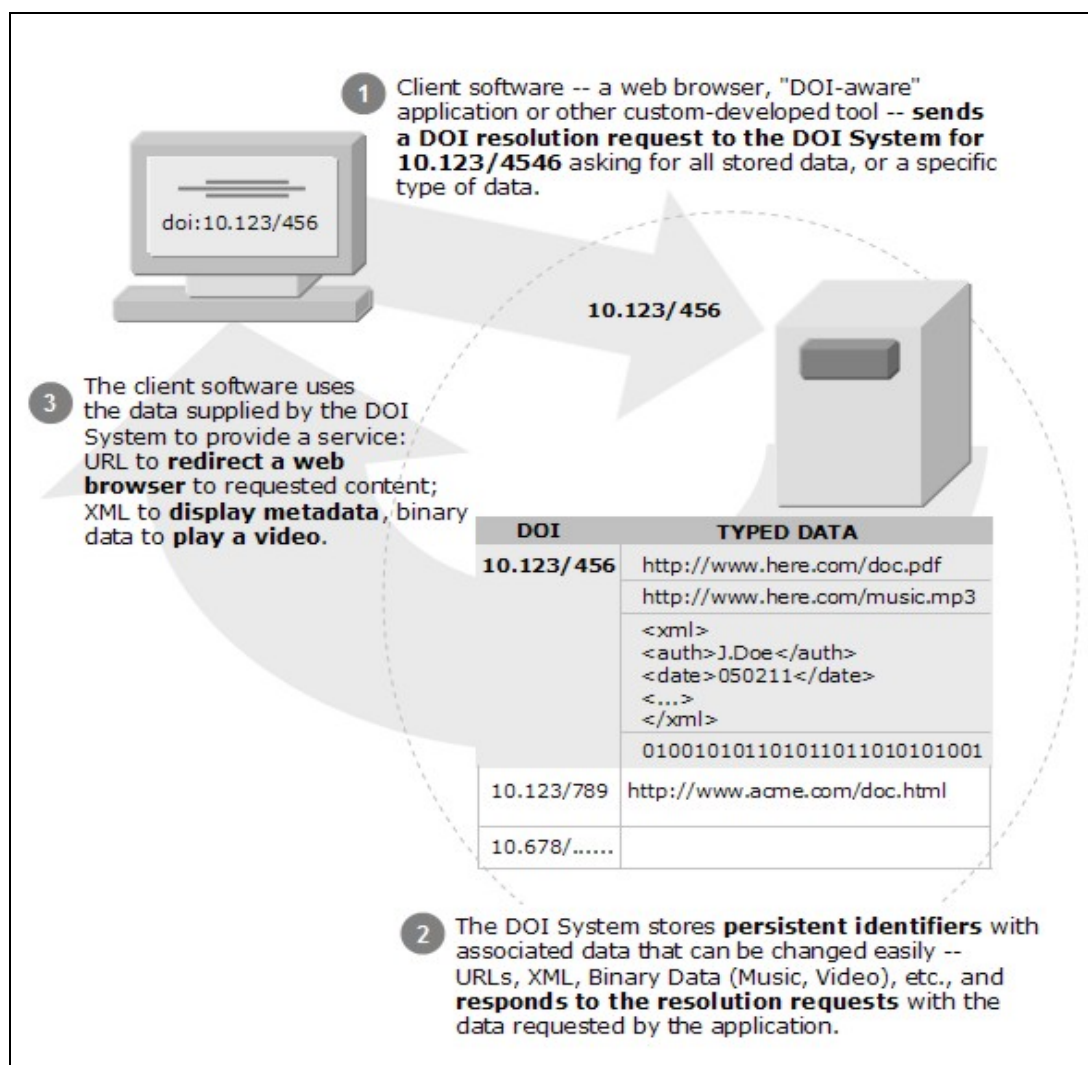
Obr. č. 5/9: Metadatový záznam disertace z institucionálního repozitáře systém DSpace Knihoven MIT s trvalým identifikátorem HDL [2007-11-30]

5.5.5 DOI

Identifikátor digitálního objektu **DOI** (Digital Object Identifier) je trvalým identifikátorem, jehož ucelený systém vznikl na základě existujících nebo rozvíjejících se standardů, zejména pak standardů Systému Handle pro směřování jmen a dále standardů projektu <indecs> (interoperability of data in e-commerce systems), který specifikoval popisná metadata [PASKIN, 2007]. DOI byl vytvořen pro potřeby jednoznačné identifikace a komunikace **digitálních objektů** (digitálních zdrojů, digitálních entit) včetně informací o nich (metadat), a to zejména v prostředí sítě Internet. DOI je využíván hojně zejména ve sféře komerčních nakladatelů. Digitální objekt může mít libovolný obsah: může nést informace textové, zvukové, obrazové a jiné povahy nebo může zahrnovat jejich kombinace.

Identifikátor DOI je možné (i v souladu s dokumentací DOI) přiřadit různým bibliografickým entitám, především **vyjádřením tvůrčích děl** (originální digitální text ve zdrojovém formátu, digitální záznam koncertu, digitální záznam divadelního představení, digitální soubor ve formátu VRML apod.) a **zhmotněním děl, tj. publikacím** (článek v časopise, digitální kniha, digitální zvukový záznam na CD-ROM atd.). DOI lze přidělit také **abstrakcím typu tvůrčí dílo** (uplatní se ale pouze v příslušném metadatovém záznamu) V prostředí sítě Internet je identifikátor DOI schopen sjednocovat stejné digitální objekty, vyskytující se na různých

místech Internetu (s různými adresami URL), v repozitářích digitálních knihoven (s lokálními digitálními identifikátory) nebo databázích (s lokálními identifikátory). Schéma na obr. č. 5/10 znázorňuje proces **směrování** digitálního objektu s určitou URL adresou přes identifikátor DOI (10.123/456) ze strany uživatele užívajícího webový prohlížeč.



Obr. č. 5/10: Schéma směrovací služby systému DOI (The Handle Net) na základě identifikátoru DOI [převzato z WWW dokumentace, 2007-11-30]

Reálné směrování identifikátorů DOI k objektům je možné vyzkoušet buď na URL <http://dx.doi.org/> nebo <http://www.handle.net/> (v obou případech jde o proxy server, který zajišťuje americká centrála systému The Handle Net).

Systém DOI byl původně od roku 1996 rozvíjen a řízen Americkou asociací vydavatelů AAP (American Association of Publishers) s podporou CNRI a vydavatelstvím Bowker. Od roku 1998 je plně řízen **Mezinárodní nadací DOI** (International DOI Foundation, <http://www.doi.org/>).

Cílem systému DOI je jeho globální rozšíření s tím, že identifikátor DOI s návaznými metadaty by se staly univerzálním prostředkem identifikace v kybernetickém prostoru. Dílčím cílem bylo zachování kompatibility a možnost využití jiných, již existujících identifikátorů

(ISBN, ISSN, amerického identifikátoru SICI, URL, URN, URI aj.). Systém DOI má značný význam pro oblast elektronického obchodu a správu autorských práv v digitálním prostředí

Syntax (struktura) identifikátoru DOI vychází plně ze syntaxe (struktury) identifikátoru HDL (Handle). V současné době je dána národní normou USA Z39.84 z roku 2000 [ANSI/NISO Z39.84-2000]:

Na návrh Mezinárodní nadace DOI se také připravuje mezinárodní norma ISO 26324 pro identifikátor DOI a jeho celý systém [ISO 26324:200?].

Informace o aktuálním stavu projektu pro zpracování mezinárodní normy jsou na URL:

<http://www.collectionscanada.ca/iso/tc46sc9/wg7/index.html>

Identifikátor DOI se skládá ze 2 základních částí oddělených lomítkem. Identifikátor může zahrnovat libovolné znaky podle Univerzálního souboru znaků UCS-2 (je definován souborem Unicode, verze 2.0). Zapisuje se v následujícím formátu:

10.1000/182

10.1145/336597.336660

10.1016/j.acalib.2006.04.004

V ukázkách jde o reálně přidělené identifikátory DOI. V prvním příkladě je identifikátor DOI přidělený manuálu Mezinárodní nadace DOI [cit přímo: International DOI Foundation, 2006], ve druhém DOI přidělený stati z konferenčního sborníku, který publikovala společnost ACM [cit. přímo: SNYMAN, 2000] a ve třetím příkladě DOI přidělený článku z časopisu *Journal of Academic librarianship* [cit přímo: COYLE, 2006].

První základní část identifikátoru DOI se, stejně jako u identifikátoru Handle, nazývá **prefix** a je složena také ze dvou základních vnitřních částí oddělených povinnou tečkou. První část prefixu se nazývá **Indikátor Adresáře** (Directory indicator) a v současné době obsahuje jen číslo „10“ (viz schéma na obr. č. 7 nebo příklady výše v textu), které reprezentuje systém DOI (všechny identifikátory systému DOI začínají tímto číslem, odlišují se tak od jiných implementací systému „The Handle System“). Druhou částí prefixu (je oddělena syntaktickou tečkou) je řetězec číslic (kód) reprezentující **registrující instituci** (Registrant code). Jde o instituci či organizaci, která si přeje registrovat digitální objekt. DOI zatím užívá jen číslice, ale v budoucnu není vyloučeno ani užívání písmen. Kód registrující organizace je možné podle stávajícího standardu a další dokumentace systému DOI doplnit dalšími dílčími prefixy (oddělovaných také tečkami), jak je vidět na následující ukázce (jde o nereálný DOI):

10.1000.10/123456

Druhá základní část identifikátoru DOI se nazývá **suffix**. Jde o jedinečné pořadové číslo, které identifikuje danou bibliografickou entitu. Přiděluje ho organizace zaregistrovaná v systému DOI. Syntax a sémantiku suffixu si organizace určuje sama. Suffix může být jednoduchým číselným řetězcem (bez významu), může však také zahrnovat písmena. V další ukázce je v suffixu aplikován jedinečný identifikátor ISBN:

10.100X/ISBN-900512-44-0

Celý systém DOI je spravován a řízen **Mezinárodní nadací DOI (IDF)**. IDF je v roli organizace, která má, kromě jiného, na starosti také správu nad dílčími **registračními agenturami DOI** (RA, Registration Agencies). Registrační agentury přidělují prefixy jednotlivým institucím, které přidělují identifikátory DOI svým informačním zdrojům. K registračním agenturám v současnosti patří například (společnosti, vydavatelé aj.):

- CrossRef (<http://www.crossref.org/>)

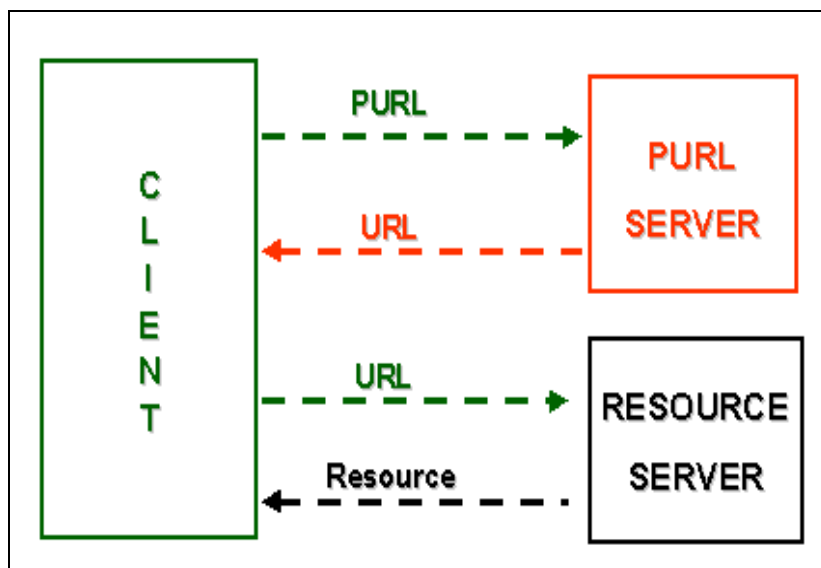
- Nielsen BookData (<http://www.nielsenbookdata.co.uk/>)
- MEDRA (<http://www.medra.org/>)
- OPOCE (Office des publications EU, <http://publications.europa.eu/>)
- R.R. Bowker (<http://www.bowker.com/>)
- TIB (Technische Informationsbibliothek) (<http://www.tib-hannover.de/>)

Bezesporu nejvýznamnější registrační agenturou systému DOI je v současnosti komerčně dostupný systém **CrossRef** (<http://www.crossref.org/>). Do provozu byl uveden v roce 2000 neziskovou a nezávislou Asociací pro mezinárodní propojování vydavatelů (Publishers International Linking Association, PILA). Jde o síť kolaborativního charakteru, jejímž hlavním posláním je **propojování online vědeckých publikací** na základě **propojování bibliografických záznamů** včetně **záznamů citací**. Založením této sítě její tvůrci reagovali zejména na **problém neplatných hyperlinků při odkazování na dokumenty lokalizované v síti Internet**. Systém CrossRef byl původně vytvořen s cílem **zlepšit přístup** k materiálům z oblasti přírodních a technických věd a medicíny, ale postupně byl rozšířen na všechny obory. Systém samotný nezahrnuje databáze úplných elektronických textů a neposkytuje propojovací služby přímo koncovým uživatelům. Umožňuje účastníkům se vydavatelům sdílet metadata dokumentů a poskytuje jejich online službám podporu při **propojování záznamů**. Výsledkem je vysoce účinný propojovací systém, jehož prostřednictvím vědec-uživatel může kliknout na libovolnou citaci v online časopiseckém článku a během okamžiku tento citovaný dokument v elektronické formě získá.

Identifikátor DOI je významným nástrojem služby CrossRef. V rámci systému CrossRef je DOI přidělován digitálním objektům (článkům, kapitolám nebo i celým dokumentům). V systému CrossRef je spojován se základními metadaty a s URL identifikovaného objektu. CrossRef je oficiální agenturou pro přidělování identifikátorů DOI. Každému digitálnímu dokumentu, který má být v systému CrossRef registrován, vytváří příslušný vydavatel kód DOI (identifikace vydavatele je uvedena v jeho prefixu). Kódy DOI jsou propojovány s **metadaty (včetně abstraktu, citací a URL)**, která tvoří vydavatelé. Ti je v definovaném formátu XML ukládají ve specifické databázi metadat **MDDB** systému CrossRef. Párová informace DOI+URL je pak následně automatickým procesem transferována do centrálního adresáře DOI s tím, že se vytváří trvalá vazba mezi DOI a dokumenty lokalizovanými v repozitářích vydavatelských digitálních knihoven. V rámci vyhledávacích procesů je využívána **směrovací služba Mezinárodní nadace DOI** (<http://dx.doi.org/>). Systém CrossRef integroval a využívá také identifikátorů **OpenURL**. Uživatel, který v rámci vyhledávání sleduje hyperlink DOI, vyvolá automatický proces vyhledání metadat z databáze systému CrossRef a následné vytvoření OpenURL, které je pak zasláno do lokální směrovací služby. Úplný text hledaného článku pak může být získán z různých zdrojů v rámci knihovnických služeb.

5.5.6 PURL

Trvalý identifikátor URL (Persistent URL, PURL) byl jednou z prvních implementací trvalých identifikátorů založených na specifikaci URN. Byl vyvinut OCLC jako jmenná a směrovací služba pro internetové zdroje, které OCLC zpracovává v rámci souborného katalogu. V roce 1996 byl identifikátor PURL nasazen v rámci Projektu internetové katalogizace (Internet Cataloguing Project) [HILSE, 2006, s. 32].



Obr. č. 5/11: schéma směrovacího systému identifikátoru PURL (připraveno podle [SHAFER, 1996])

Identifikátor PURL může identifikovat jak tištěné, tak elektronické dokumenty. Jde o **trvalé identifikátory** informačních zdrojů, ale také o **lokátory**, které odkazují na bezprostřední směrovací službu. Identifikátor PURL je v podstatě identifikátorem URL s tím, že PURL nevede přímo k internetovému zdroji, ale ke směrovací službě (viz obr. č. 5/11). Záznamy v databázi serveru směrovací služby (PURL server) obsahují identifikátor PURL a jeden nebo více odpovídajících URL, na kterých se daný informační zdroj nachází. Pro směrování se využívá i v tomto případě standardního protokolu HTTP. Informační zdroj lze pak získat prostřednictvím klientského prohlížeče, který automaticky zajistí získání jeho digitální kopie.

Hlavními cíli OCLC v projektu PURL bylo:

- Oddělení lokátorů (adres URL) od jmen internetových zdrojů
- Využití standardních služeb a protokolů
- Trvalost, která je zajištěna softwarem OCLC (měnit lze URL, nikoliv trvalý identifikátor PURL)

Identifikátory PURL se skládají ze tří částí:

Příklady:

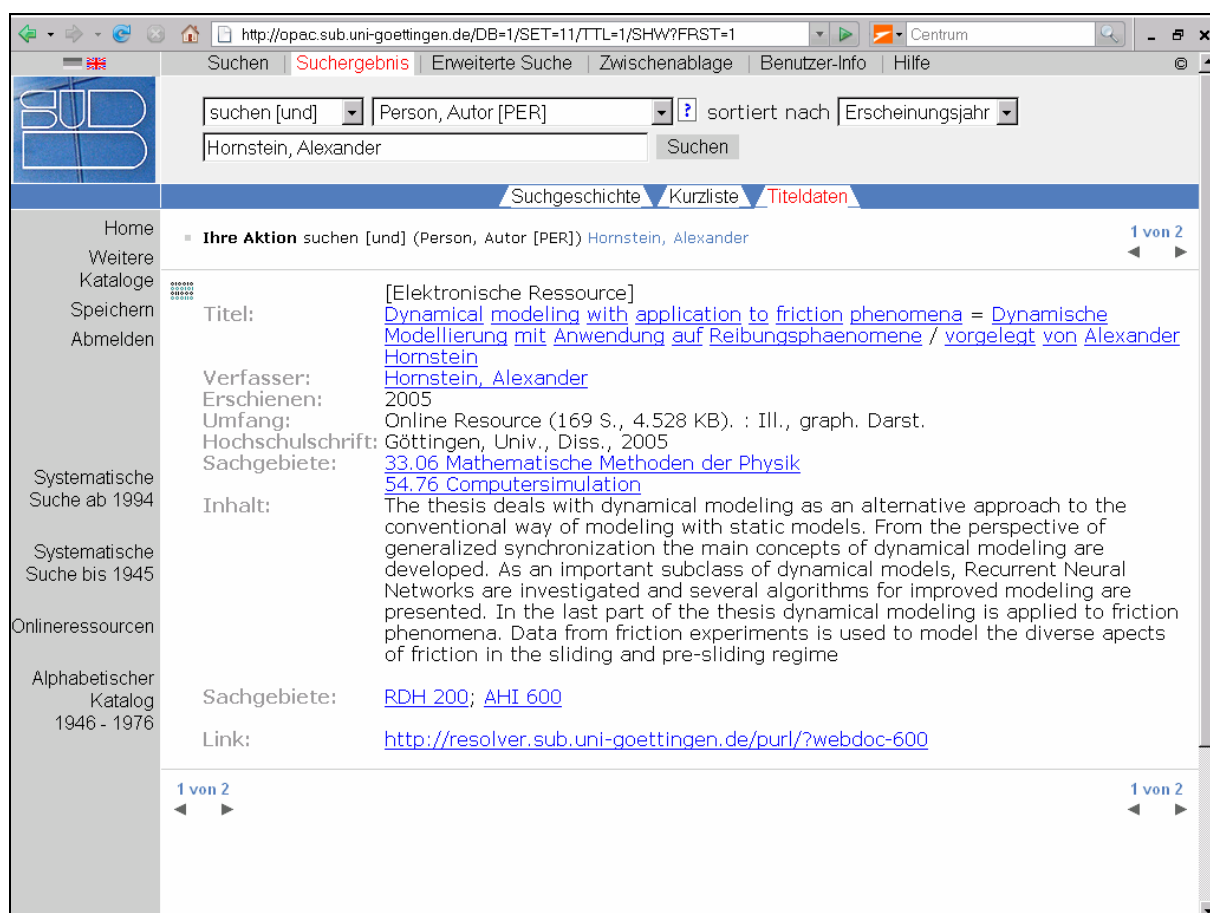
<http://purl.oclc.org/keith/home>

<http://purl.oclc.org/OCLC/PURL/FAQ>

První částí je označení **protokolu** (http), druhou služba doménového jména (DNS), která zajišťuje přístup ke směrovací službě přes příslušnou IP adresu, a třetí částí je jméno informačního zdroje. Software pro PURL server, vytvořený OCLC, je volně dostupný z webového sídla systému PURL (<http://purl.oclc.org>). Software zahrnuje vše, co je potřebné pro údržbu identifikátorů PURL v distribuovaném prostředí.

Na obr. č. 5/12 je zobrazen katalogizační záznam disertační práce Univerzitní knihovny Göttingenské univerzity (OPAC systému PICARTA), který zahrnuje také identifikátor PURL (viz návštěví Link:):

<http://resolver.sub.uni-goettingen.de/purl/?webdoc-600>



Obr. č. 5/12: Záznam disertace z katalogu Univerzitní knihovny Göttingenské univerzity ve výstupním formátu systému PICARTA s trvalým identifikátorem PURL [2007-11-30]

Po kliknutí je uživatel směrován z identifikátoru PURL prostřednictvím směrovací služby a webového prohlížeče na metadatový záznam, který je lokalizován na dokumentačním serveru Univerzitní knihovny (viz obr. č. 5/13). Reálné URL metadatového záznamu je:

<http://webdoc.sub.gwdg.de/diss/2005/hornstein/>

V metadatovém záznamu uživatel najde odkaz na úplný text, který se nachází na reálném URL:

<http://webdoc.sub.gwdg.de/diss/2005/hornstein/hornstein.pdf>

Pokud uživatel vyhledá záznam stejné disertace v katalogu Německé národní knihovny (text je uložen na jejím depozitním serveru jako povinný výtisk), zjistí také URN této disertace, jehož specifický řetězec je tvořen modifikovaným jménem zdroje z identifikátoru PURL:

<urn:nbn:de:gbv:7-webdoc-600-8>

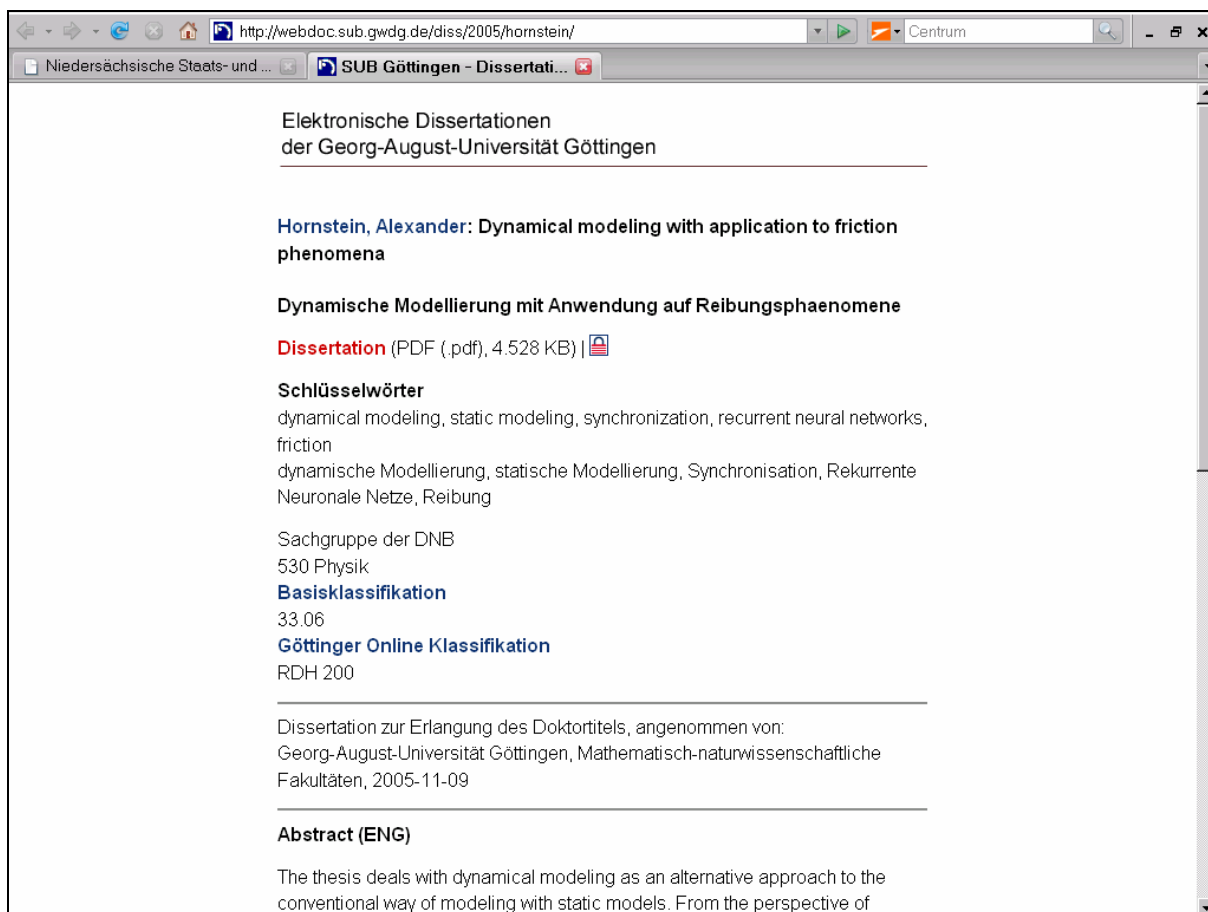
(jméno metadatového záznamu v identifikátoru PURL bylo „webdoc-600“, jméno dokumentu je „webdoc-600-8“).

Směrování z daného URN bude vedeno přes směrovací službu NNK:

<http://nbn-resolving.de/urn:nbn:de:gbv:7-webdoc-600-8>

přímo k adrese URL úplného textu:

<http://webdoc.sub.gwdg.de/diss/2005/hornstein/hornstein.pdf>



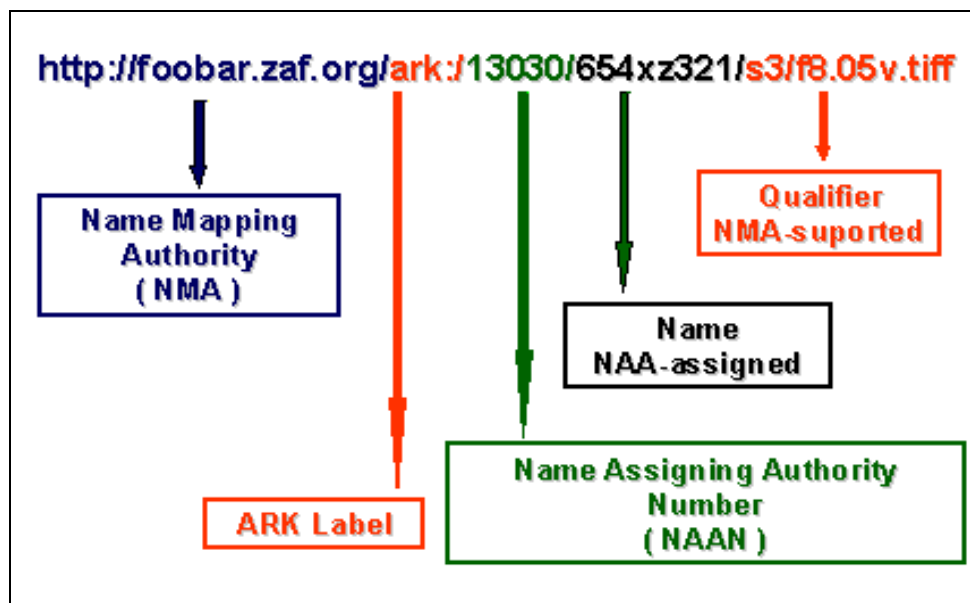
Obr. č. 5/13: Metadatový záznam disertace z repozitáře dokumentačního serveru Göttingenské univerzity s odkazem na úplný text [2007-11-30]

5.5.7 AKR

Identifikátor **ARK** (Archival Resource Key, čes. Klíč archivního zdroje) náleží k novějším trvalým identifikátorům, které byly vyvinuty zejména pro potřeby identifikace digitálních zdrojů. Má progresivní koncepci, která je příslibem do budoucnosti. Jde v podstatě o **specifické URL** vytvořené za účelem trvalého a dlouhodobého přístupu k informačnímu zdroji (objektu). Identifikátory ARK jsou, nicméně, určeny k identifikaci informačních zdrojů libovolného typu. Vedle **digitálních zdrojů** (digitálních dokumentů, obrazů, elektronických databází, softwaru, webových sídel aj.) mohou také identifikovat **fyzické objekty** (tradiční knihy, sochy, archeologické nálezy) nebo **nehmatatelné objekty** (chemické látky, nemoce, slovníkové termíny, umělecké výkony apod.). Cílem tvůrců identifikátoru ARK bylo vytvořit stabilní „jméno“, resp. **odkaz (referenci)**, který by byl **trvale spojen** s informačním objektem (zdrojem) [KUNZE, 2007].

Identifikátor ARK byl vyvinut Johnem A. Kunzem (z Kalifornské univerzity), a to v rámci přípravy studie o systémech trvalých identifikátorů, zpracovávané pro Národní lékařskou knihovnu USA (NLM). Text byl zveřejněn jako Internetový návrh poprvé v roce 2001, aktuální verze je z května 2008 [KUNZE, 2008]:

Schéma identifikátoru ARK je rozvíjeno v rámci dokumentace Kalifornské digitální knihovny (CDL, <http://www.cdlib.org>).



Obr. č. 5/15: schéma struktury identifikátoru ARK (připraveno podle [KUNZE, 2007])

Identifikátor ARK zahrnuje dvě základní části (viz obr. č. 5/15). Hlavní částí je **stabilní a neměnné jméno informačního zdroje** (je uvozeno návěštím-zkratkou „ark:“). Hlavní části předchází **nepovinná a flexibilní část**, která se může stát předmětem změny (bude obsahovat jméno serveru, který zajišťuje potřebné služby).

V pořadí první část identifikátoru (celého „URL“) zahrnuje protokol a jméno serveru, který bude mapovat jméno zdroje (Name Mapping Authority, NMA). Jde o „**akční**“ (v rámci webového prohlížeče klikatelnou) část celého identifikátoru. (V aktuálně platné verzi je uvedeno modifikované pojmenování serveru - „Name Mapping Authority Hostport (NMAH)“).

Druhá (hlavní) část je uvozena návěštím včetně syntaktické dvojtečky - „arc:“. Následuje schválené a registrované jméno instituce či organizace, která přiděluje zbytek kódu celého identifikátoru ARK (Name Assigning Authority, NAA; zkratka NAAN znamená číslo této instituce či organizace). Hodnota uvedená na obr. č. 11 - „13330“ patří Kalifornské digitální knihovně.

V současné době je již přidělena řada dalších jmen (v podobě číselné hodnoty), která jsou v seznamu vedena na serveru Kalifornské univerzity. Vybrané příklady jmen:

NAAN	NAA
12025	National Library of Medicine
12026	Library of Congress
12027	National Library of Agriculture
12148	Bibliothèque nationale de France (Gallica)
13030	California Digital Library
13038	World Intellectual Property Organization
20775	University of California SanDiego
28722	University of California Berkeley
29114	University of California San Francisco
62624	New York University Libraries
78428	University of Washington
80444	Northwest Digital Archive
88435	Princeton University

Jméno informačního objektu (Name - viz obr. č. 5/15) a popřípadě i jeho kvalifikátor (Qualifier) lze generovat pomocí libovolného generátoru (softwaru), který bude vyhovovat specifikacím identifikátoru ARK. K dispozici je také software „**noid**“ (**nice opaque identifiers**), který je možné získat zdarma po registraci na Kalifornské univerzitě.

Na základě identifikátoru ARK lze realizovat řadu služeb. Klikáním na celý identifikátor v prostředí WWW bude uživatel propojen k více objektům najednou (nikoli jen k jednomu), tj.:

- K objektu samotnému nebo jeho dílčí částem
- K metadatovému záznamu
- K fasetovému vystavení všech dílčích objektů zahrnutých pod jeden identifikátor ARK (vhodné je za tímto účelem využití prohlížeče Mozilla Firefox)

Názorný příklad aplikace identifikátoru ARK je dále v textu demonstrován na digitalizované souborné publikaci *Současná francouzská literatura XIX. století* (6-ti svazkové dílo autora J.M. Quérarda aj., publikované v Paříži v letech 1842-1856), která je uložena v digitální knihovně Národní knihovny Francie zvané **Gallica** (<http://gallica.bnf.fr>). Trvalý identifikátor ARC byl v tomto digitálním systému nasazen v nedávné době [BERMÈS, 2006]. Gallica po registraci získala v rámci systému ARC kód „12148“.

Výchozí **identifikátor ARC** pro množinu všech objektů prvního svazku celého díla se zobrazí v pozici adresy URL v rámci webového prohlížeče po vyhledání citovaného titulu v rámci rešeršního rozhraní systému Gallica a výběru 1. svazku celého díla. Je následující:

<http://gallica.bnf.fr/ark:/12148/bpt6k209395s>

Jde o objekt **prezentující celý první neskenovaný svazek** (stranu po straně ve zmenšeném formátu s možností listování), pro jeho pojmenování přímo v systému Gallica je užito francouzského odborného termínu pro vydání publikace „chemin de fer“. Úplné znění identifikátoru ARK pro tento objekt zahrnuje syntaktickou tečku umístěnou za základ identifikátoru a termín „cheminde fer“:

<http://gallica.bnf.fr/ark:/12148/bpt6k209395s.chemindefer>

Ze standardní zobrazené stránky se může uživatel přemístit k záznamu prvního svazku citované publikace (viz obr. č. 5/16), jehož identifikátor ARK je:

<http://gallica.bnf.fr/ark:/12148/bpt6k209395s.notice>

Zobrazit lze i objekt se stránkováním prvního svazku díla

<http://gallica.bnf.fr/ark:/12148/bpt6k209395s.pagination>

nebo zobrazení celého dokumentu prvního svazku díla:

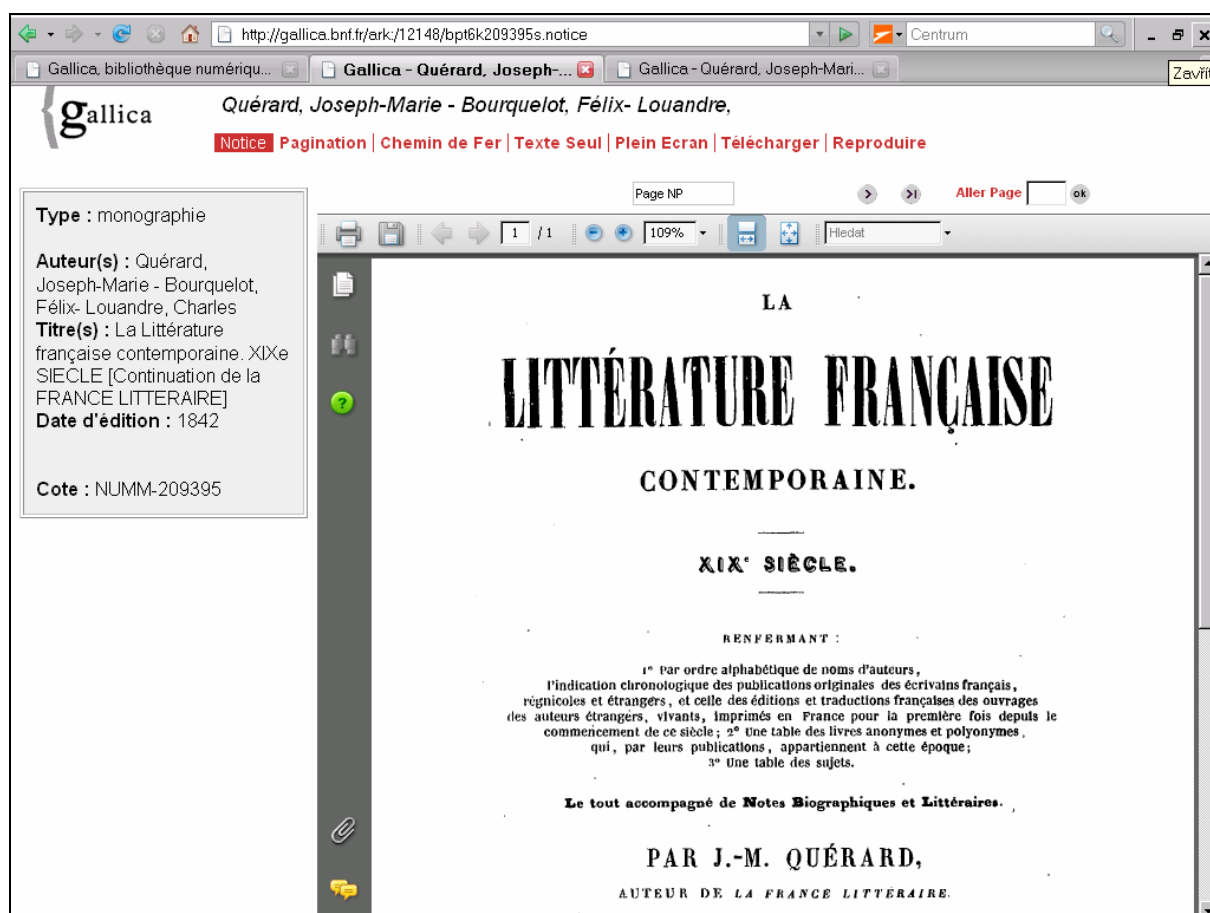
<http://gallica.bnf.fr/ark:/12148/bpt6k209395s.item>

Pro srovnání: kmenový identifikátor ARK pro druhý svazek díla (Tom 2) je:

<http://gallica.bnf.fr/ark:/12148/bpt6k2093918>

a kmenový identifikátor ARK pro třetí svazek díla (Tom 3)

<http://gallica.bnf.fr/ark:/12148/bpt6k2093918>



Obr. č. 5/16: záznam digitalizovaného díla (vlevo) s identifikátorem ARK (v adrese prohlížeče) části náhledu na titulní stranu v digitální knihovně Národní knihovny Francie Gallica [2007-11-22]

5.6 Trvalá identifikace fyzických osob a korporací v prostředí digitálních archivů : aktuální směry řešení v zahraničních systémech

5.6.1 Úvod do problematiky - sledovaný problém

Pracovní skupina Odborné komise pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací (<http://www.evskp.cz/>) AKVŠ, která v první polovině roku 2008 připravovala ostrou verzi (1.1) národního standardu *Metadatový soubor pro popis elektronických vysokoškolských kvalifikačních prací v ČR (EVSKP-MS)* [EVSKP-MS, 2008-07-15] a souvisejících specifikací pro popis fyzických osob *PersCZ* (verze 1.0) [PersCZ, 2008] a pro popis korporací *CorpCZ* (verze 1.0), se nutně musela zabývat také otázkou **jednoznačné a globálně trvalé identifikace fyzických osob**, které vystupují v roli autorů, vedoucích nebo oponentů vysokoškolských prací (viz prvky „dc:creator“ a „dc:contributor“ ve standardu *EVSKP-MS*), a také **korporací**, které vystupují v roli vydavatelů a archivujících institucí těchto prací nebo institucí, které přidělují příslušné tituly (tamtéž, viz prvky „dc:publisher“ a „thesis:grantor“).

Za provizorní a prozatím za ne zcela jednoznačné identifikování autoritních entit lze považovat doporučení standardu připojovat (v jednodušším případě a nepovinně) ke jménu autora práce doplňující informace typu rok narození, získané tituly apod. [EVSKP-MS, 2008-07-15, dc:creator]. U vedoucích a oponentů závěrečných prací je ve standardu nabídnuta možnost připojení alespoň identifikačního čísla autoritního záznamu z národní databáze *AUT*

(<http://sigma.nkp.cz/F/>) [EVSKP-MS, 2008-07-15, dc:contributor], kterou spravuje Národní knihovna ČR, byť z globálního hlediska prozatím nejde o identifikátor jedinečný. Z hlediska technologického však zatím není možné hypertextové propojování záznamů vysokoškolských prací uložených v lokálních databázích škol s autoritním záznamem v národní databázi *AUT*. Problémem také je, že ne každý vysokoškolský pedagog nebo oponent autoritní záznam včetně jeho identifikátoru v národní databázi *AUT* má. Výjimečně by bylo možné využít čísla autoritního záznamu také k jednoznačnější identifikaci autorů vysokoškolských prací s podmínkou, že již publikovali a v bázi *AUT* mají reprezentaci (zejména v případě doktorandů). To se týká také studentů, kteří obhájili a publikovali disertaci v zahraničí. Pokud je taková publikace získána do fondů našich knihoven, autoritní záznam se v národní databázi objeví (viz například záznam s hodnotou identifikačního čísla „jx20050601022“ v poli 001 a také v poli 100, podpoli |7 ve formátu MARC 21).

U korporací typu vysokých škol a jejich součástí (včetně knihoven) se s ohledem na jejich neveliké množství v ČR lze prozatím spoléhat na vytvoření autorizovaného seznamu jejich jmen s případným doplněním jednoznačných, ale zatím jen lokálně platných identifikačních kódů s omezením na ČR, jakými jsou například identifikátor školy RID v databázi REGPRO matriky SIMS aj. [EVSKP-MS, 2008-07-15, dc:publisher, evskp:contact]. Využít by bylo možné i v tomto případě identifikačního čísla záznamu korporace z databáze *AUT* (opět prozatím bez hypertextového propojení mimo systém NK ČR).

Oba navazující standardy pro popis fyzických osob *PersCZ* a pro popis korporací *CorpCZ* zahrnují vnořené svěbytné prvky pro zápis dat narození, titulů aj., nicméně jejich hodnoty nemohou ani v tomto případě řešit trvalou a globálně jednoznačnou identifikaci příslušných autoritních entit. Pro budoucí využití byl již nyní do standardů uveden (zatím nepovinný) prvek „**dc:identifier**“, který je určen pro libovolný **trvalý a globálně jedinečný identifikátor**. Někteří autoři si již takové identifikátory zavedli v zahraničních systémech, nicméně zavedení a rutinní provoz dosazování trvalých identifikátorů všem autorům vysokoškolských prací však bude záviset na vývoji systémů pro správu takových identifikátorů v samotné ČR. V současnosti takový systém v ČR zaveden není.

Dotazy členů skupiny komise vznesené na odborníky NK ČR spravující databázi *AUT* a také jednání s nimi počátkem roku 2008 ohledně možnosti vytváření autoritních záznamů a využívání identifikačních čísel těchto záznamů k identifikaci autorů vysokoškolských prací prozatím nevedlo k žádnému k reálnému řešení. V databázi *AUT* se v současné době v rámci stanovené politiky nevytvářejí autoritní záznamy na autory vysokoškolských prací, pokud nejsou publikačně činní. Nabídka NK ČR, aby si vysoké školy takové autoritní záznamy vytvářely samy s tím, že by bylo možné jejich zařazování do databáze *AUT*, je zvažována, realizace této nabídky však v současné chvíli naráží na řadu objektivních organizačních, systémových, finančních, personálních a jiných potíží.

Otázka, která pracovní skupinu připravující národní standardy zajímá, je, jak se řeší a realizuje trvalá a globálně jedinečná identifikace autoritních entit v zahraničních systémech. Následující text je stručným přehledem vybraných přístupů a výsledků řešení této problematiky, a to v oblasti knihovnické, v oblasti komerčních informačních systémů a také v oblasti digitálních archivů.

5.6.2 Identifikace autoritních entit v knihovnické oblasti

Knihovníci jsou tou komunitou, která se otázkou jedinečné identifikace autorů zabývá nejdéle, v podstatě od té doby, co se začaly katalogizovat knihy v rámci příslušných katalogizačních zvyklostí a pravidel. Několik posledních desetiletí národní bibliografické agentury v řadě zemí spravují a nadále rozvíjejí **soubory autoritních záznamů**. V prvním

období šlo o papírové formy, v současnosti již běžně o formy počítačových databází autoritních záznamů, které jsou zpravidla vždy propojeny se záznamy katalogizačních databází uvnitř provozů příslušných knihoven (národních aj.). Jak správně upozornila Amanda Hillová, knihovny však ve svých souborech autoritních záznamů nezahrnují **autory článků z časopisů**, kterých jen v oblasti vědy a výzkumu existuje obrovské množství, a nezahrnují v současnosti ani neustále narůstající velkou množinu **autorů preprintů, výzkumných zpráv, vysokoškolských prací a jiné šedé literatury**, kteří tyto dokumenty ukládají do centralizovaných předmětově vymezených archivů nebo institucionálních repozitářů [HILL, 2008, Introduction]. Jednoznačné identifikování autorů v narůstajícím množství různorodých informačních systémů se však jeví jako nutnost a tuto nutnost pocítují také sami vědci, kteří se v tomto roce často i v odborném tisku ptali po možnosti zavedení trvalých identifikátorů, které by jim pomohly například lépe sledovat citovanost vlastních děl [AERTS, 2008b, s. 979; AERTS, 2008a, s. 575; KAPPENBERG, 2008].

Knihovnická komunita se vedle péče o problematiku souborů autoritních záznamů zabývá již řadu let otázkou **trvalých identifikátorů autoritních entit** (fyzických osob, korporací aj.), jejich povahou a podstatou a také otázkou jasnějšího vymezení entit, kterým by měly být takové identifikátory přidělovány. Mezinárodní standard z této aktivity prozatím z různých příčin (finančních, organizačních aj.) nevzešel, v posledním období se však díky pokročilým informačním technologiím daří experimentovat v oblasti vytěžování dat z katalogizačních databází a budování nových webových aplikací nad nimi a dále také v oblasti kooperativního budování mezinárodních souborů **autoritních záznamů**, ve kterých hrají důležitou roli **jejich identifikační čísla**.

5.6.2.1 Plánované identifikátory ISADN a INSAN

Myšlenka o přidělování mezinárodního standardního čísla, které by jednoznačně identifikovalo **autoritní entity** v rámci autoritních záznamů (nikoliv záznamy samotné), vznikla na půdě organizace IFLA již v 70. letech 20. století. Koncepce vytvoření a zavedení Mezinárodního standardního čísla pro autoritní data **ISADN** (International Standard Authority Data Number) byla později rozvíjena v 80. letech. Návrh tohoto čísla se objevil i ve významné instrukci *Průvodce pro autoritní záhlaví a odkazy* již v roce 1984. Novější vydání této instrukce s pozměněným názvem *Průvodce pro autoritní záznamy a odkazy* (GARR, Guidelines for Authority Records and References) číslo ISADN zahrnuje také [Mezinárodní federace, 2001].

Jeden z existujících modelů mezinárodního čísla **ISADN** byl navržen francouzskými odborníky v 80. letech 20. stol. [BOURDON, 1993, s. 81]. Číslo mělo obsahovat celkem 13 znaků (arabských číslic a písmen latinské abecedy). Mělo se skládat ze čtyř částí reprezentujících určité vlastnosti autoritních entit samotných: zdroje čísla autoritní entity a zároveň autoritního záznamu (2 znaky pro zemi a 2 pro korporaci přidělující číslo), dále údaje o **národnosti/zemi** autoritní entity (2 znaky), **jazyka** užívaného v dílech entity (3 znaky) a **pořadového čísla** autoritního záznamu přiděleného v rámci organizace (4 znaky), která by záznam vytvořila [WILLER, 1996]. Příklad:

ISADN frBNfrfre2506

Na přelomu tisíciletí se objevila nová americká modifikace identifikačního čísla autoritních entit - **INSAN** (International Standard Author Number). Šlo o rozšíření výše uvedeného francouzského modelu [SNYMAN, 2000; TILLET, 2000]. Navrhovaná struktura obsahovala některé nové prvky, které měly i některé identifikátory bibliografických entit (například ISRC). Číslo INSAN mělo zahrnovat celkem 18 znaků. 2 abecední znaky byly určeny pro **kód agentury** přidělující číslo a tvořící záznam, 2 znaky pro **kód národnosti/země**, který by

reprezentoval národnost fyzické osoby nebo zemi sídla korporace, 3 znaky by byly určeny pro **kód jazyka**, v němž se objevuje převážná část publikací dané entity, 4 číselné znaky pro **rok přidělení čísla**, 6 číslic by reprezentovalo jednoduché **pořadové číslo** autoritního záznamu, který by přidělovala organizace tvořící záznam, a poslední číslo mělo být **kontrolní číslicí**.
Příklad:

INSAN LCcaeng2003002288-5

Uvedené modely identifikátorů nebyly v knihovnické praxi nikdy zavedeny. Nicméně IFLA a její nová **Pracovní skupina pro Funkční požadavky a číslování autoritních záznamů FRANAR** (Working Group on Functional Requirements and Numbering of Authority Records), která připravila již druhou verzi koncepčního modelu s definicí funkčních požadavků na autoritní záznamy [Mezinárodní federace, 2005; Mezinárodní federace, 2007], studuje i nadále proveditelnost a realizaci čísla ISADN, nikoliv však v původní koncepci. Zároveň skupina bedlivě sleduje práce jiných zainteresovaných skupin, které se zabývají realizací trvalých identifikátorů autoritních entit v jiných oblastech, včetně skupiny, která připravuje mezinárodní normu **identifikátoru ISNI** [ISO Project 27729]. Podle názoru členky pracovní skupiny FRANAR Barbary Tilletové výhody jedinečných čísel k identifikaci autoritních entit sice existují, jsou však převažovány velkými náklady, které by musely být vynaloženy k vybudování globálního systému pro jejich přidělování [TILLET, 2007, s. 343, s. 359]. Autorka vyjadřuje názor, že je možné budovat jiné aplikace, které budou těžit z dat z existujících knihovnických databází. Také v poslední zprávě skupiny FRANAR, kterou připravila B. Tilletová, je v závěrech zopakován diskutabilní názor, že identifikačního čísla ISADN pro autoritní entity samotné není zapotřebí [TILLET, 2008, s. 1, s. 12].

V rámci diskusí odborné skupiny FRANAR byla zároveň rozvedena otázka předmětu jedinečného identifikování. Každá z oblastí či komunit (knihovny, archivy, muzea, vydavatelé, autorsko-právní organizace aj.) v současnosti chápe entity k identifikování (identity) podle svých vlastních potřeb. Knihovníci upřednostňují pojetí osoby jako „**bibliografické identity**“ (angl. „persona“), to znamená v podstatě entitu, která se skrývá za **personálním jménem (nebo více personálními jmény)**, bez ohledu na to, jde-li o reálnou osobu nebo nikoliv. Například známý francouzský romanopisec s vlastním jménem Romain Kacew (1914-1980), který publikoval díla pod více pseudonymy, je v národním souboru autoritních dat USA reprezentován 2 autoritními záznamy, jeden reprezentuje pseudonym Émile Ajar (viz autoritní záznam <http://errol.oclc.org/laf/n2006-53657.html>), druhý pseudonym Romain Gary (viz autoritní záznam <http://errol.oclc.org/laf/n50-16999.html>). (Zajímavostí je, že v národním francouzském souboru autoritních záznamů je ustanovena pouze jedna forma jména jako preferovaná - pseudonym Romain Gary (viz autoritní záznam s trvalým identifikátorem ARK (<http://catalogue.bnf.fr/ark:/12148/cb11904228r/>)). Autorsko-právní organizace nebo jejich mezinárodní federace potřebují spíše identifikaci reálné osoby, neboť ve hře jsou honoráře za využívání autorských děl [TILLET, 2007, s. 345].

5.6.2.2 Projekt „WorldCat Identities“ (OCLC) a služba LAF (LC a OCLC)

Projekt „*WorldCat Identities*“ (Identity Světového katalogu) je novou informační službou (aplikací) výzkumného oddělení OCLC v oblasti jednoznačné identifikace fyzických osob a korporací s připojením řady návazných informací (seznam děl dané osoby, seznam publikací o dané osobě, časový rámec publikační aktivity, alternativní formy jména, propojení na autoritní záznam pocházející z Kongresové knihovny, propojení na Wikipedii aj.). Služba má zatím experimentální ráz (Beta verze) a je v současné době lokalizována na jednom na serverů OCLC (OAICat, který je provozován v rámci výzkumu ALCME - Advanced Library

Collection Management Environment). Stávající server (<http://orlabs.oclc.org/identities/>) má být v nejbližším termínu změněn.

Původním cílem této služby bylo uvedení úplného přehledu všech autorů (fyzických osob i korporací), kteří jsou uvedeni v záznamech databáze Světového katalogu OCLC a nasměrovat tak uživatele zpět k využití katalogizačních záznamů a potažmo fondů všech zastoupených knihoven. Pracovníci výzkumného oddělení však zjistili, že služba má možnosti dalšího širšího využití [DANSKIN, 2008]. Perspektivně například právě v oblasti jednoznačné identifikace autoritních entit, a to i přes to, že uživatel v současné chvíli naráží často na různé problémy nesprávné identifikace osob a jejich děl, určitě pak v případech, kdy není připojen záznam z databáze autoritních záznamů Kongresové knihovny (LC NAF, Library of Congress Name Authority File).

Data jsou sklízena z databáze Světového katalogu pomocí protokolu OAI-PMH a jsou ukládána do databáze na serveru **OAICat**. V současné době služba nabízí na cca **15 000 000-20 000 000 jmen** (přibližně 500 000 jmen je uvedeno v nelatinském písmu). Protože katalog zahrnuje také záznamy vysokoškolských prací (především autorů z území USA), jsou v rámci služby „*WorldCat Identities*“ prezentovány i záznamy těchto autorů, většinou ovšem bez návazného propojení na autoritní záznam pocházející z Kongresové knihovny (LC NAF). Autoritní záznam je vytvářen jen tehdy, když autor již něco publikoval a knihovny dané publikace mají ve fondech - viz například záznam bývalého českého doktoranda na obr. č. 5/17 a obr. č. 5/18.

The screenshot displays the 'WorldCat Identities' web application. At the top, there is a search bar with the text 'Search Identities' and a 'Name:' input field followed by a 'Go' button. Below the search bar, a navigation menu includes links for 'Jump To: Overview | Publication Timeline | Works About | Works By | Audience Level | Related Names | Useful Links | Fast Headings'. The main heading is 'Sobotka, Tomáš'. To the right, under the 'Languages' section, it shows 'English (4)' and 'No Linguistic content (2)'. The 'Overview' section provides statistics: 'Works: 4 works in 6 publications in 2 languages and 27 library holdings' and 'Classifications: hb991,'. The 'Publication Timeline' section features a bar chart from 2000 to 2010, with a key indicating orange bars for 'Publications about Tomáš Sobotka' and red bars for 'Publications by Tomáš Sobotka'. The 'Most widely held works by Tomáš Sobotka' section lists four titles: 1. 'Postponement of childbearing and low fertility in Europe' (2 editions, 2004, 16 libraries); 2. 'Ten years of rapid fertility changes in the European post-communist countries : evidence and interpretation' (2 editions, 2002, 8 libraries); 3. 'Missing births" : decomposing the declining number of births in Europe into tempo, quantum and age structure effects' (1 edition, 2005, 1 library); 4. 'Religiosity and spatial demographic differences in the Netherlands' (1 edition, 2002, 1 library). The browser's status bar at the bottom shows 'Internet' and '100%' zoom.





Obr. č. 5/17: část záznamu českého autora (též autora disertace) z databáze služby „*WorldCat Identities*“ (OCLC) [převzato z WWW 2008-10-30]

V rámci služby „*WorldCat Identities*“ se využívá, když existuje, **jedinečného identifikačního čísla autoritního záznamu ze souboru LC NAF**. Například identifikační číslo autoritního záznamu autora, prezentovaného na obr. č. 2, má hodnotu:

no2002-94899

Toto číslo (předchází mu zkratka „lccn-“, která označuje výchozí zdroj čísla) se tudíž objevuje jako součást jedinečného identifikátoru záznamu typu **trvalého URL** (Persistent URL) v databázi služby „*WorldCat Identities*“ (<http://orlabs.oclc.org/identities/lccn-no2002-94899>). Na této adrese lze také záznam v současné době na WWW bez potíží nalézt.

V rámci uživatelského rozhraní služby „*WorldCat Identities*“ je možné vyhledávat následující typy záznamů autoritních entit (při vyhledávání je využito grafických ikon):

	Personální identita (A Personal Identity) - modrá ikona
	Personální identita z řízeného slovníku (A Personal Identity from a controlled vocabulary) - oranžová ikona
	Korporativní identita (A Corporate Identity) - modrá ikona
	Korporativní identita z řízeného slovníku (A Corporate Identity from a controlled vocabulary) - oranžová ikona

OCLC ONLINE COMPUTER LIBRARY CENTER
A Project of OCLC Research

LAF
Linked Authority File

Identifier: no2002-94899
Persistent URL: <http://errol.oclc.org/laf/no2002-94899.html>
XML Record: <http://errol.oclc.org/laf/no2002-94899.MarcXML>

0000000nz 2200000n 0000
001oca05889089
00520021018052755.0
008021017n| acannaab |n aaa c
010|ano2002094899
040|aNjP|beng|cNjP
1001 |aSobotka, Tomáš
670|aTen years of rapid fertility changes in the European post-communist countries, 2002:|bt.p.
(Tomáš Sobotka)
999|a28

These authority records are derived from the [Library of Congress authority file](#). The purpose is to support [interactive](#) and [automated](#) authority lookups and provide Web access to individual authority records.

Obr. č. 5/18: autoritní záznam včetně trvalého URL českého autora (též autora disertace) ve formátu MARC 21 ze služby LAF (OCLC) [převzato z WWW 2008-10-30]

Po základní rešerši systém zpravidla zobrazuje seznam jmen, jejichž text bývá zobrazován v různé velikosti (včetně ikon) - jde o symbolické znázornění relativní popularity identit.

Propojování záznamů služby „*WorldCat Identities*“ na autoritní záznamy, pokud existují, se uskutečňuje na základě existence jiné databáze, kterou připravilo v rámci výzkumu OCLC (vedoucím projektu je Jeff Young). V rámci experimentu jsou z výchozí databáze autoritních záznamů USA (LC NAF), kterou spravuje Kongresová knihovna, sklizeny opět pomocí protokolu OAI-PMH všechny její autoritní záznamy. Záznamy jsou takto sekundárně k dispozici v nové WWW aplikaci nazvané „*LAF (Linked Authority File)*“ (<http://alcme.oclc.org/laf/index.html>). Zobrazovat je lze jak ve formátu MARC 21, tak ve formátu MarcXML. Jedinečné číslo autoritního záznamu je i v této aplikaci součástí identifikátoru typu trvalého URL (Persistent URL). Již výše uvedené identifikační číslo autoritního záznamu s hodnotou „no2002-94899“ (viz obr. č. 5/18) se tudíž také objevuje jako součást jedinečného identifikátoru autoritního záznamu typu **trvalého URL** (Persistent URL) v databázi služby „*LAF (Linked Authority File)*“ (<http://errol.oclc.org/laf/no2002-94899.html> nebo <http://errol.oclc.org/laf/no2002-94899.MarcXML>). Hodnota identifikátoru je v autoritním záznamu uvedena ve formátu MARC 21 v poli 010 (viz obr. č. 5/18).

5.6.2.3 Projekt Virtuálního mezinárodního autoritního souboru (VIAF)

V předcházející části zmíněný server OCLC (<http://orlabs.oclc.org/>) je hostitelským serverem také pro společný mezinárodní projekt označovaný zkratkou **VIAF** (Virtual International Authority File, (<http://orlabs.oclc.org/viaf/>), který má za úkol propojit mezi sebou zatím vybrané národní soubory autoritních záznamů. K dispozici je také jeho virtuální adresa (<http://viaf.org/>). V současné době se projektu účastní Kongresová knihovna ve Washingtonu (<http://www.loc.gov/>) se svým souborem **LC NAF** (LC Name Authority File), Německá národní knihovna (<http://www.d-nb.de/>) se svým souborem **PND** (Personennormdatei) a nově také Národní knihovna Francie (<http://www.bnf.fr/>) se svým vlastním souborem autoritních záznamů (Fichiers d'autorité auteurs-titres) a také OCLC (<http://www.oclc.org/>). V budoucnu by se měly připojit také jiné národní knihovny (včetně Národní knihovny ČR).

V rámci projektu VIAF je vytvářen virtuální mezinárodní soubor autoritních záznamů na základě spojení dat z národních souborů. Výsledným efektem je zachování národních variant forem jmen autoritních entit, které jsou hypertextově propojovány mezi sebou pomocí identifikačních čísel. V rámci jednoduchého rozhraní lze vyhledávat jména osob s tím, že lze získat buď jen jeden nebo více autoritních záznamů najednou (národní agentury jednotlivých zemí totiž vytvářejí autoritní záznamy do určité míry multiplicitně). Na obr. č. 5/19 je zobrazen autoritní záznam stejného českého autora jako na obrázku č. 3 (s identifikátorem no2002-94899) ve formátu MARC 21 OCLC, nalezený přes rozhraní systému VIAF (záznam pochází z výchozího souboru LC NAF). Záznam lze za příznivých podmínek fungování sítě Internet získat z URL adresy: <http://orlabs.oclc.org/viaf/LC%7Cno2002094899>

U řady autoritních záznamů z výchozího souboru LC NAF jsou na konci záznamů připraveny hypertextové odkazy, které vedou k záznamům dokumentů v katalogu Kongresové knihovny. Propojení je připraveno na základě tzv. „**permalinků**“ (permalinks, tedy trvalých hypertextových odkazů), jimiž jsou pomocí nové aplikace opatřeny již všechny katalogizační záznamy. Při propojování se využívá nového **info URI schématu** [RFC 4452, 2006], které vzniklo na návrh knihovnické a vydavatelské komunity. Souvisí úzce s rozvojem amerického standardu OpenURL, který je jedním z podstatných nástrojů efektivního propojování různorodých informačních zdrojů. Konkrétní jmenný prostor **info:lccn/** je definován pro identifikátory záznamů pocházejících z Kongresové knihovny. Je zaregistrován v centru info URI schémat, které má na starosti OCLC (<http://info-uri.info/>).

Virtual International Authority File
Hosted by OCLC Research

Names
Sobotka, Tomáš
Search

LC Record for Sobotka, Tomáš

Fixed	Rec stat	n	Entered	2002-10-17	Replaced	2002-10-18T05:27:55.0
	Type	z	Upd status	a	Enc lvl	Source c
	Roman		Ref status	n	Mod rec	Name use a
	Govt agn		Auth status	a	Subj a	Subj use a
	Series	n	Auth/ref	a	Geo subd n	Ser use b
	Ser num	n	Name	a	Subdiv tp	Rules c

001	LC no2002094899
010	no2002094899
035	(OCoLC)oca05889089
040	NjP ‡b eng ‡c NjP
100	1 Sobotka, Tomáš
670	Ten years of rapid fertility changes in the European post-communist countries, 2002: ‡b t.p. (Tomáš Sobotka)
919	ten years of rapid fertility changes in the european post communist countries ‡9 1
943	200x ‡9 1

Obr. č. 5/19: autoritní záznam českého autora (též autora disertace) ve formátu MARC 21 získaný z aplikace VIAF (OCLC) [převzato z WWW 2008-10-06]

5.6.2.4 Trvalý identifikátor ARK pro záznamy autoritních entit

Identifikátor **ARK** (Archival Resource Key, čes. Klíč archivního zdroje) patří k novějším trvalým identifikátorům, které byly vyvinuty zejména pro potřeby identifikace digitálních zdrojů. Má progresivní koncepci, která je příslibem do budoucnosti. Jde v podstatě o **specifické URL** vytvořené za účelem trvalého a dlouhodobého přístupu k informačnímu zdroji (objektu) [KUNZE, 2008]. Identifikátory ARK jsou určeny také k identifikaci autoritních entit. Toho bohatě využila Národní knihovna Francie (BnF), která tento identifikátor nasadila nejen do svojí digitální knihovny Gallica pro potřeby trvalé identifikace samotných digitalizovaných knih, ale také pro potřeby trvalé identifikace katalogizačních záznamů a také záznamů autoritních.

Například významný autor 17. století Gabriel Naudé (1600-1653) je v souboru autoritních záznamů reprezentován záznamem s trvalým identifikátorem:

ark:/12148/cb11917444g

Směrování tohoto identifikátoru je zajištěno přes směrovací systém (server) Národní knihovny Francie pomocí URL adresy:

<http://catalogue.bnf.fr/ark:/12148/cb11917444g>

Celkově systém zajišťuje tři formáty zobrazení autoritního záznamu:

- ve formátu pro veřejnost
(<http://catalogue.bnf.fr/ark:/12148/cb11917444g/PUBLIC>)
- ve výměnném národním formátu INTERMARC
(<http://catalogue.bnf.fr/ark:/12148/cb11917444g/INTERMARC>)
- a ve formátu UNIMARC-A
(<http://catalogue.bnf.fr/ark:/12148/cb11917444g/UNIMARC>)

Základ trvalého identifikátoru je ve všech případech stejný (ark:/12148/cb11917444g).

Notice d'autorité personne

Rappel de la recherche :
IDP=cb12044263p

Mes recherches | Mes préférences | Réservations | Panier

Affichage public | Intermarc | Unimarc

Voir les notices

000 cx a22 45
001FRBNF120442639
00519880419
009<http://catalogue.bnf.fr/ark:/12148/cb12044263p>
039 \$oOPL\$a003429567
100 \$a19870706afrey0103 ba0
101 \$afre
102 \$aFR
103 \$a 19510323
105 \$aa
106 \$a1
152 \$aAFNOR Z 44-061
200 |\$7ba0yba0y\$8fre \$9 0 \$aBourdon\$bFrançoise\$f1951-....
3000 \$aConservateur à la Bibliothèque nationale (depuis 1982)
801 \$aFR\$bBnF\$c19880419
810 \$aLa Bibliographie spécialisée en sciences humaines au
CAFB / F. Bourdon, G. Maxit, 1984
<http://catalogue.bnf.fr/ark:/12148/cb12044263p/UNIMARC>

Obr. č. 5/20: autoritní záznam s trvalým identifikátorem ARK (ark:/12148/cb12044263p) francouzské autorky ve formátu UNIMARC z francouzského národního souboru autoritních záznamů [převzato z WWW 2008-10-06]

Celý identifikátor ARK zahrnuje dvě základní části. Hlavní částí je **stabilní a neměnné jméno informačního zdroje** (je uvozeno návěštím „ark:“). Hlavní částí předchází **nepovinná a flexibilní část**, která se může stát předmětem změny (obsahuje protokol a jméno serveru, který zajišťuje potřebné služby). Tato první část identifikátoru (celého „URL“) mapuje jméno zdroje (Name Mapping Authority, NMA). Jde o „akční“ (v rámci WWW prohlížeče klikatelnou) část celého identifikátoru. Druhá (hlavní) část je po lomítku uvozena návěštím včetně syntaktické dvojtečky - „ark:“. Po lomítku následuje schválené a registrované jméno instituce, která přiděluje zbytek kódu celého identifikátoru ARK (Name Assigning Authority, NAA). Hodnota uvedená výše v příkladech „12148“ patří Národní knihovně Francie. Poslední částí identifikátoru je vlastní identifikátor s případným nepovinným kvalifikátorem (například označení formátu, jako PUBLIC, INTERMARC a UNIMARC).

Výše uvedený autor Gabriel Naudé je autorem příručkového díla k budování knihovny *Advis pour dresser une bibliothèque*, jejíž druhé vydání z roku 1644 bylo zdigitalizováno a je uloženo v digitální knihovně Gallica (stávající první verzi systému) s identifikátorem ARK (<http://gallica.bnf.fr/ark:/12148/bpt6k6514x>) nebo (v nové druhé verzi systému) se stejným základním identifikátorem (<http://gallica2.bnf.fr/ark:/12148/bpt6k6514x>).

Na obr. č. 5/20 je zobrazen autoritní záznam ve formátu UNIMARC(A) současné francouzské autorky, která se věnuje intenzivně oblasti autoritní kontroly včetně oblasti trvalých identifikátorů, Françoise Bourdonové. Nad záznamem (Rappel de la recherche) je vlastní jméno identifikátoru ARK (IDP=cb12044263p), celý identifikátor ARC je pak uložen v poli 009 (<http://catalogue.bnf.fr/ark:/12148/cb12044263p/UNIMARC>), extra je zobrazen ještě pod záznamem. Záhlaví této autorky v seznamu použité literatury [BOURDON, 1993] je hypertextově propojeno na její autoritní záznam pomocí trvalého identifikátoru ARK.

5.6.3 Identifikátory v oblasti autorských ochranných svazů

5.6.3.1 Identifikátor IPI (IPIN) v sektoru správy autorských práv

Identifikátor IPI (IPIN) pochází z **oblasti komunity společností pro správu autorských práv**. Je doposud užívaný v rámci systému CIS (Common Information System), který spravuje Mezinárodní konfederace CISAC (<http://www.cisac.org>). Identifikátor tzv. **zainteresovaných stran IPI** (Interested Parties Information, popř. IPIN Interested Parties Information Number) je globálním identifikátorem **fyzických osob a korporací**, jež jsou v roli tvůrců děl (skladatelů, textařů, interpretů, režisérů, spisovatelů aj.) nebo majitelů jednotlivých typů práv (autorských a jiných) k těmto dílům. Identifikátor **IPI** je vytvářen a přidělován na základě interních metodik. Číslo **IPI** (popř. **IPIN**) je přidělováno centrem IPI v rámci budování databáze informací o zainteresovaných stranách **IPI**, která je pro CIS a CISAC provozována švýcarskou společností pro správu autorských práv hudebních děl **SUISA** (Swiss Society for the Rights of Authors of Musical Works, <http://www.suisa.ch/>). Databáze IPI je online dostupná, ale zatím jenom pro oprávněné uživatele přes WWW (<http://www.ipisystem.org/>).

Všichni členové společností sdružených v rámci Mezinárodní konfederace CISAC mají číslo IPI (IPIN) přidělováno automaticky. Jedinečný **identifikátor IPI** zahrnuje celkem 13 znaků a skládá se z prefixu I, vlastního identifikátoru a kontrolního čísla. Příklad (z veřejné dokumentace CISAC):

I-000 000 229-7

V databázi, která v současnosti obsahuje cca 3 000 000 záznamů, je identifikátor doplňován údajem o jménu tvůrce či majitele práv, údaji o datu narození, národnosti, pseudonymech aj.

5.6.3.2 Mezinárodní identifikátor veřejných identit ISNI (projekt ISO normy)

V dubnu 2006 podala Mezinárodní konfederace CISAC návrh na novou mezinárodní normu s označením a číslem **ISO 27729 ISPI** (International Standard Party Identifier). Předmětem návrhu měla být mezinárodní identifikace všech „stran“ (entit) zainteresovaných na tvorbě, výrobě a distribuci tvůrčích děl [ISO Project 27729, 2006]. Trvalý identifikátor měl významně přispět k potřebné interoperabilitě v automatizovaných systémech nebo digitálních aplikacích napříč různorodými oblastmi (oblast vydavatelů, distributorů, informačních institucí, knihoven apod.). Přípravou normy je pověřena pracovní skupina (WG 6, TC 46/SC 9, ISO), která zahrnuje řadu expertů z různých oblastí, informačních systémů, členy skupiny se stali také experti z francouzské, britské a finské národní knihovny.

V roce 2007 se řešila řada koncepčních otázek [ISO Project 27729, 2007]. Velká diskuse byla věnována problému **vymezení identifikované entity** a vymezení množiny metadat tuto entitu popisujících. Debata vyústila k návrhu na změnu označení identifikátoru z ISPI na **ISNI** (International Standard Name Identifier) [LLORET, 2007b, s. 4]. Někteří odborníci poukázali na skutečnost, že v návrhu normy nebyla ujasněna otázka identifikovaných entit, v textu se vyskytovaly termíny, jako „**fyzická a korporativní jména**“ (natural and corporate names), „**fyzická nebo právnická osoba**“ (natural or legal person), „**tvůrce**“ (creator) a „**fyzická osoba nebo právnická entita**“ (natural person or corporate entity). Hlavní původní termín „**strana**“ (party) nebyl jasně definován. Zástupci knihoven při té příležitosti zároveň poukázali na koncepční model FRAD (Funkční požadavky na autoritní data), který byl ve druhé verzi publikován organizací IFLA v dubnu 2007 [Mezinárodní federace, 2007]. Jména a identifikátory jsou totiž v modelu FRAD pojímána jako oddělené entity, jež jsou odlišné od entit, na které se autoritní záznam zaměřuje. Poukázáno bylo na problematiku pseudonymů, kdy může existovat více jmen, které patří jediné fyzické osobě (problém utajování osob z různých důvodů a obecně otázka veřejné identity osob) [LLORET, 2007a, s. 4].

Diskuse vyústila do nového modifikovaného vymezení hlavních pojmů pracovní komisí WG 6 (bylo schváleno v květnu 2007). Připomenuta byla **veřejná identita** identifikovaných entit. Pojem „**strana**“ (party) bude v navržené normě chápán jako veřejně orientovaný (public oriented) [LLORET, 2007b, s. 3]:

- **Strana** je fyzická nebo právnická osoba nebo skupina fyzických nebo právnických osob, které jsou zahrnuty do tvorby nebo rozšiřování intelektuálního vlastnictví
- **Veřejná identita** je identita, která je spojena a veřejně užívaná nějakou stranou; většina reálných osob má jednu jedinou veřejnou identitu, některé jich ale mohou mít víc
- **Identifikátor ISPI [nově ISNI]** bude identifikátorem **veřejné identity** nějaké strany

Systém pro přidělování identifikátoru ISNI prozatím neexistuje. Předpokládá se vytvoření **Mezinárodní agentury ISNI**, která bude spravovat celý globální systém. Systém bude zahrnovat řadu **registračních agentur ISNI** vymezených na národní či jiné úrovni. Některé subjekty, které značně pokročily v oblasti identifikace osob, se již nyní o tuto roli dopředu ucházejí - například nizozemská OCLC-PICA by se ráda po dohodě s nadací SURF této role ujala za Nizozemsko [SPANJE, 2007, s. 11]. Registrační agentury budou přidělovat jedinečné **kódy ISNI** na základě žádosti registrující osoby (autora, skladatele, vydavatele, knihovníka nebo jiné právnické osoby). Přidělení ISNI bude vázáno na registraci v příslušné registrační databázi, která bude zahrnovat také **metadata** (povinná a volitelná). Podrobné pokyny k registraci bude schvalovat a vydávat Mezinárodní agentura ISNI.

Návrh nového standardu **ISO 27729 ISNI** předpokládá **13-místný kód** pro jednoznačnou a trvalou identifikaci autoritních entit („stran“), které jsou v roli tvůrců, producentů, výrobců či distributorů děl [ISO Project 27729, 2006]. Číslo by mělo být **nevýznamové** a mělo by mít pravděpodobně následující strukturu: písmenný prefix ISNI identifikátor registrační agentury, identifikátor autoritní entity a kontrolní číslo:

ISPI 999 999 999 999-7

Navržený model může v průběhu přípravy normy zaznamenat modifikace.

Mezinárodní norma prochází posledními fázemi její přípravy, dle posledních sdělení pracovní skupiny musí být finální návrh schválen nejpozději do 2. února 2009 a její zveřejnění musí nastat nejpozději v srpnu 2009 [Terms of Reference, 2007].

5.6.4 Trvalý identifikátor „ResearcherID“ (komerční sféra)

Americká společnost Thomson Scientific (<http://scientific.thomson.com>, část společnosti Thomson Corporation), která produkuje a komerčně zpřístupňuje jedinečný vědecký citační systém *SCI (Science Citation Index)* prostřednictvím portálu *Web Of Knowledge, WOK - Web of Science, WOS*), uvedla v lednu 2008 do provozu zcela novou, specifickou webovou službu nazvanou anglicky „**ResearcherID**“ (Identifikátor vědce, <http://www.researcherid.com/>). Cílem této služby je registrace vědců či badatelů (též doktorandů), v jejímž rámci je přidělován **jedinečný globálně trvalý identifikátor** (Researcher ID number). Globální jedinečnost základního čísla, resp. kódu, je dána vytvořením stabilní internetové domény. Jde sice o službu komerčního charakteru, která je prioritně určena vědcům, kteří jsou předplatiteli systému WOK (k této službě se hlásí přímo z rozhraní tohoto systému), nicméně o registraci se mohou ucházet i další zájemci, pokud o to požádají a od služby obdrží pozvánku (v tomto případě neproběhne přidělení vstupních práv ihned, ale zhruba do týdne po ručním zpracování žádosti) [Thomson Scientific, 2008, Registration]. Odpovědní pracovníci dbají o kvalitu členství. Registrace není nijak vázána na podmínku mít registrované záznamy článků v bázi systému WOK/WOS, je to však výhodou, neboť nová aplikace je automaticky připravena na veřejné zobrazování záznamů zdrojových článků daného vědce z této databáze s návaznými citačními službami, které jsou ovšem dostupné již jen uvnitř editačního rozhraní. Vědec může uvnitř systému dle předepsané metodiky a předepsaného formátu připojovat i záznamy jiných dokumentů, které publikoval, aniž by byly registrovány v databázi WOK/WOS.

Služba „ResearcherID“ je perspektivním příspěvkem k řešení letitého problému jednoznačné identifikace autorů v citačním rejstříku *SCI* (v budoucnu se předpokládá využívání této aplikace k řízení jmen autorů i v rámci systému WOK/WOS). Identifikátor je možné využívat i v jiných systémech či službách, neboť je volně dostupný včetně metadatového záznamu. Jde o moderní webovou aplikaci, která umožňuje, aby si vědec sám vyplnil a průběžně aktualizoval určitou sadu personálních údajů, k nimž patří (viz obr. č. 5/21) preferovaná forma příjmení a křestního jména, jiné formy jména (včetně variant s diakritikou apod.), emailová adresa, pracoviště vědce, vědecký profil vyjádřený pomocí předmětové kategorie, klíčových slov a popisu. Uživatel uvnitř editačního rozhraní sám rozhoduje, které údaje se, kromě údajů standardních, budou zobrazovat veřejně a které ne. V rámci registračního procesu je vědci přidělen identifikátor, který je v rámci záznamu zobrazován také jako ucelená adresa typu trvalého URL, například (viz též obr. č. 5/21, kde je zobrazen záznam bývalého doktoranda z ČR s afiliací platnou v září 2008):

<http://www.researcherid.com/rid/B-7276-2008>

Základnímu kódu předchází jméno adresáře (rid/). Kód zahrnuje na začátku písmeno (např. A, B, C), lze předpokládat, že jde o určitou kategorizaci (ve veřejně dostupné dokumentaci producenta není kategorizace osvětlena [Thomson Scientific, 2008, What is a ResearcherID number]). Po spojovníku je uvedeno registrační číslo vědce a na konci je uveden rok přidělení kódu. Další příklady přidělených kódů jsou uvedeny v závěrečném seznamu citovaných pramenů - jsou skryty pod hyperodkazem u jmen autorů [AERTS, 2008; HILL, 2008, aj.].

Aplikace nabízí libovolným návštěvníkům z internetu možnost volného vyhledávání záznamů podle vybraných selekčních údajů (identifikátoru, jména autora, afiliace, klíčových slov aj.). Připraven je také hypertextově propojený rejstřík klíčových slov, takže rešerši je možné zahájit i tímto způsobem.

Registrační databáze záznamů vědců, která se od ledna 2008 rychle naplňuje, by měla být podle plánů odpovědných pracovníků společnosti Thomson Scientific v nejbližší době zajišťována sponzory z okruhu univerzit, vydavatelů vědeckých asociací aj. [HILL, 2008, s. 28]. Další plány společnosti jsou ambiciózní, v budoucích verzích aplikace je plánováno

propojování registračních záznamů „ResearcherID“ do jiných sociálních sítí, blogů, různých databázových aplikací (viz například španělský adresář expertů v oblasti zpracování informací EXIT, <http://www.directorioexit.info/>), který tento identifikátor v rámci záznamů registruje, existuje-li pro danou osobu), propojování na texty prací vědců pomocí identifikátoru DOI aj.

ResearcherID
Researcherid.com

Home Login Search

Researcher Profile **Sramek, Ondrej** Return to Search Page Preview the New Features Labs

ResearcherID: **B-7276-2008** Last / Family Name: **Sramek** First / Given Name: **Ondrej**

Other Names: **Ondřej Šrámek (accented spelling)** Institution: **Charles University in Prague, Faculty of Mathematics and Physics**

E-mail: **sramek@geo.mff.cuni.cz** Sub-org / Dept: **Department of Geophysics**

URL: **<http://www.researcherid.com/rid/B-7276-2008>** Address: **V Holešovičkách 2**

Role: **Researcher (Academic)** Address: **Praha 8**

Subject: **Geochemistry & Geophysics** Address: **180 00**

Keywords: **geophysics ; geophysical fluid dynamics ; multiphase flow** Address: **Czech Republic**

Description: **Dynamics and early evolution of Earth and planets, core formation, partial melting and melt migration.**

Publication List Citation Metrics

1 publication(s) Page 1 of 1 Go Sort by: Times Cited

1. Title: **Simultaneous melting and compaction in deformable two-phase media**
Author(s): **SRAMEK, O; RICARD, Y; BERCOVICI, D**
Source: **GEOPHYSICAL JOURNAL INTERNATIONAL** Volume: **168** Issue: **3** Pages: **964-982** Year: **MAR 2007**
Times Cited: **1**
DOI: **10.1111/j.1365-246X.2006.03269.x**

1 publication(s) Page 1 of 1 Go Sort by: Times Cited

Obr. č. 5/21: metadatový záznam českého doktoranda s trvalým identifikátorem lokalizovaný ve webové aplikaci ResearcherID (Thomson Scientific) [převzato z WWW 2008-09-30]

5.6.5 Identifikátory v oblasti digitálních repozitářů

5.6.5.1 Trvalý Digitální identifikátor autora (Nizozemsko)

Jedním z nejzajímavějších projektů, které v posledních letech úspěšně řešily otázku trvalé identifikace autorů digitálních obsahů, je nizozemský projekt **Digitálního identifikátoru autora DAI** (Digital Author Identifier, <http://dai-uitrol.ub.rug.nl/>). Vlastním cílem projektu bylo přidělit **jedinečné identifikační číslo** všem aktivním badatelům (včetně doktorandů), kteří působí v Nizozemsku. Účelem bylo vytvoření spolehlivých propojení informací o badatelích a vědeckých skupin (korporací) s informacemi o jejich vědecké produkci (publikované i nepublikované literatuře, datových souborech a jiných informačních zdrojích). Systém přidělování čísel zahrnoval vybudování specifické databáze autoritních dat.

Projekt **DAI** se původně začal rozvíjet v rámci širšího nizozemského **programu DARE** (Digital Academic Repositories), jenž vznikl na základě iniciativy řady akademických a výzkumných institucí v Nizozemsku (zahrnoval 13 nizozemských univerzit, Královskou

nizozemskou akademii pro umění a vědu KNAW, Národní organizaci výzkumných organizací NWO, Královskou knihovnu aj.). Program DARE, který byl rozvíjen v letech 2003-2006 pod patronací a koordinací nadace SURF (<http://www.surffoundation.nl/>), vedl ke **vzniku sítě digitálních akademických repozitářů** v Nizozemsku **DAREnet** včetně jeho původního samostatného portálu (<http://www.darenet.nl/>). Od 1. 1. 2007 spravuje celou síť Oddělení pro výzkumné informace akademie KNAW (<http://www.knaw.nl/>). V současné době je původní portál již součástí Národního informačního systému pro akademický výzkum a spolupráci **NARCIS** (National Academic Research and Collaboration Information System). Jeho nová portálová aplikace (<http://www.narcis.info/>) je v současnosti již také funkční a poskytuje volně informace o cca **40 000** badatelích, údaje o přibližně **163 000** vědeckých publikacích (volně dostupných z existujících institucionálních repozitářů nizozemských univerzit), informace o asi **2 380** datových souborech pocházejících z nizozemských výzkumů, dále je možno získat přibližně **20 700** úplných textů doktorských disertací a také okolo **48 000** publikací **217** špičkových nizozemských akademiků (údaje z října 2008).

Projekt DAI vznikl nejdříve na Groningenské státní univerzitě (<http://www.rug.nl/>), později se rozšířil ve dvou fázích do dalších 12 nizozemských univerzit [SPANJE, 2007, s. 2]. Autoritní data v **databázi DAI** jsou dnes udržována společně **pracovníky knihoven** nizozemských univerzit (mají na starosti správu univerzitních institucionálních repozitářů) a **pracovníky oddělení pro vědu a výzkum** těchto univerzit. Tato oddělení jsou zároveň účastníky nizozemského administrativního informačního systému pro vědu a výzkum **METIS** (<http://metis.hosting.kun.nl/>), který v Nizozemsku funguje již od roku 1993, jeho centrum je na Radboudské univerzitě v Nijmegen (Radboud Universiteit in Nijmegen, <http://www.ru.nl/>).

Každému badateli v Nizozemsku je přidělováno **jedinečné identifikační číslo**, a to bez ohledu na to, jestli již publikoval vědecké články nebo ne. Identifikátor DAI tvoří nezbytnou součást záznamu badatele (včetně autorů obhájených disertačních prací), který je uložen v automatizovaném knihovnickém systému univerzity (v souboru autoritních záznamů). K tomuto záznamu jsou automatickou cestou přidávány specifické údaje z databází oddělení pro výzkum a vývoj (METIS) dané univerzity (například forma jména badatele z registru METIS, podrobné údaje o jeho afiliaci včetně kódů institucí, kontaktní údaje, údaje o funkcích, vědeckém profilu, historii zaměstnání badatele aj.). Databáze kompletních autoritních dat (včetně identifikátoru DAI) není veřejně dostupná a její správu zajišťuje národní knihovní síť PICA (knihovny univerzit jsou jejími účastníky).

Základem **trvalého identifikátoru DAI** se stalo identifikační číslo z národního souboru autoritních záznamů nizozemské knihovní sítě PICA. Tomuto číslu předchází nově schválený trvalý identifikátor typu URI „info“ schématu:

info:eu-repo/dai/nl

který si 13. května 2008 nechala ve světové centrále URI „info“ schémat (<http://info-uri.info/>) zaregistrovat nizozemská nadace SURF. Hlavní část schváleného URI „info“ schématu - doména **info:eu-repo/** - je připravena k využití pro libovolný digitální repozitář v Evropě. Celé URI „info“ schéma pro identifikátor DAI (**info:eu-repo/dai/**) může být aplikováno pro potřeby trvalého identifikování tvůrců obsahů v digitálních repozitářích kdekoli v Evropě. Prvními uživateli jsou tedy Nizozemci, proto je celé schéma doplněno kódem země „nl“, za kterým je již konkrétní identifikační číslo.

V rámci portálu **NARCIS** je možné vyhledávat specifické záznamy nizozemských badatelů, které v sobě v prezentační rovině obsahuje řadu podstatných informací včetně odkazu na publikační činnost. Zahrnují již také dobře viditelný trvalý identifikátor DAI. Například záznam špičkového nizozemského odborníka z oblasti statistiky prof. R. H. Koniga

(<http://www.narcis.info/person/RecordID/PRS1241113/Language/NL/>) má přidělen trvalý identifikátor DAI (viz též obr. č. 5/22):

info:eu-repo/dai/nl/07483374X

Záznam českého doktoranda T. Sobotky, který obhájil disertaci na Groningenské univerzitě, (<http://www.narcis.info/person/RecordID/PRS1285747/query/Sobotka/Language/en/#>), má přidělen trvalý identifikátor DAI:

info:eu-repo/dai/nl/240476832

The screenshot shows the NARCIS website interface. At the top, there are navigation tabs: NARCIS, DAREnet, Cream of Science, and Promise of Science. The main header area includes the text: "Cream of Science showcases prominent research from the Netherlands. The website lists the names of 217 top Dutch academics, providing worldwide access to their 47,614 publications. About 60% of these can be accessed full text. These full-text publications are a subset of NARCIS and DAREnet." There is a search bar with the word "statistics" entered and a "search" button. Below the search bar, there are checkboxes for "Organisations", "Persons" (which is checked), "Current research", "Publications (full text)", and "Data sets".

The profile for **Koning, Prof.dr. R.H.** is displayed. It includes a portrait photo of the professor. To the right of the photo is a "TagCloud" with the following tags: actuarial science, applied, econometrics, econometrics, sport, sports, economics, and statistics. The profile details include:

- Expertise:** Statistics; Applied econometrics; Actuarial science; Sport and econometrics
- Expertise (NL):** Statistiek; Toegepaste econometrie; Actuarial; Sport en econometrie
- Digital Author ID:** info:eu-repo/dai/nl/07483374X
- Active as:**
 - Extraordinary professor: Vakgroep Economics and Econometrics (RuG)
 - Chair (EN): Sports economics
 - Chair (NL): Sporteconomie
 - Phone: +31-50-3637192
 - Email: r.h.koning@rug.nl
 - URL: <http://www.rhkoning.com/>
- Publications of this Cream of Science member:**
 - Publication List
- Current research:** (the most recent research is at the top)
 - Project leader: Risk management
- Completed research projects:**
 - Completed research: Dutch Research Database (NOD)
- Classification:**
 - D11600 Probability theory, statistics

The bottom of the page shows a browser status bar with "Internet" and a 100% zoom level.

Obr. č. 5/22: záznam nizozemského vědce s trvalým digitálním identifikátorem DAI (info:eu-repo/dai/nl/07483374X) ze systému NARCIS [převzato z WWW 2008-10-30]

Identifikátory DAI prozatím nejsou směrovatelné. Na příslušné webové aplikaci se, podle sdělení Daniëla van Spanje (sít' PICA) z počátku listopadu 2008, v současnosti intenzivně pracuje. Identifikátory DAI by měly být směrovatelné z centrálního nizozemského serveru (<http://www.persistent-identifier.nl/>), který je již částečně v provozu pro potřeby směrování identifikátorů URN - ten je již přidělován například disertacím i jiným dokumentům.

Mimořádně zdařilý projekt a systém přidělování identifikátorů DAI, stejně jako celý národní nizozemský informační systém pro vědu, výzkum a spolupráci NARCIS, byly vybrány jako modelové řešení celoevropského projektu **DRIVER II** (<http://www.driver-community.eu/>).

5.6.5.2 Systém RePEc a jeho identifikátory autorů a korporací

Významný globální digitální systém pro oblast ekonomiky a příbuzných oborů **RePEc** (Research Papers in Economics, <http://repec.org/>) je jedním z největších systémů, který zpřístupňuje podstatnou část svého obsahu v digitální podobě volně (zdarma) svým uživatelům z celého světa. Decentralizovaně budovaný systém obsahuje v současnosti přibližně **630 000 dokumentů** a jejich metadatových popisů. Zhruba **520 000** z nich je dostupných online zdarma. Systém RePEc je zajišťován přibližně 100 dobrovolníků z celého světa (přibližně ze 70 zemí). Zakladatelem systému je Thomas Krichel, který má v systému jako autor dokumentů přidělen také identifikátor s pořadovým **číslem 1** (RePEc Short-ID: pkr1) (<http://ideas.repec.org/e/pkr1.html>).

Jádrem systém RePEc je několik dílčích subsystémů a služeb. Patří k nim zejména subsystém **IDEAS** (Internet Documents in Economics Access Service, <http://ideas.repec.org/>), který je kompletní bibliografickou (metadatovou) databází systému RePEc, služba **NEP** (New Economics Papers, <http://nep.repec.org/>), která zajišťuje rozšiřování informací o nových studiích z oblasti ekonomie, dále digitální repozitář **MPRA** (Munich Personal RePEc Archive, <http://mpra.repec.org/>), který slouží autorům, jež nejsou členy žádné registrované členské instituce, systém **LogEc** (<http://logec.repec.org/>) který poskytuje statistické údaje týkající se systému RePEc, služba **CitEc** (Citations in Economics, <http://citec.repec.org/>) zajišťující zpracování citací pro systém RePEc, a jiné.

Dva subsystémy zabezpečují pro RePEc jedinečné identifikační služby, podotknout je ale třeba, že jde o lokálně platné identifikátory. První z nich je **Autorská služba RePEc** (RePEc Author Service, <http://authors.repec.org/>), v jejímž rámci jsou spravovány kompletní informace o autorech systému RePEc. V současné době je v systému registrováno přibližně **18 000 autorů** (též z ČR). Službu řídí Thomas Krichel (Univerzita na Long Islandu, USA). V rámci této služby se musí registrovat každý nový autor systému RePEc. Sám vyplňuje řadu údajů přes webové rozhraní (jméno, varianty jména, afilační údaje, e-mailová adresa aj.). Systém pak pro autora zajišťuje propojování autoritního záznamu se záznamy jeho dokumentů. V rámci registračního procesu je autorům přidělen **jedinečný identifikátor**, který se skládá ze vybraných znaků jména autora a nějakého významného údaje typu datum). Krátká forma identifikátoru se pak zobrazuje také v prezentačním záznamu subsystému IDEAS. Na obr. č. 5/23 je prezentační formát záznamu autora z České republiky, doktoranda CERGE-EI, jehož zkrácený identifikátor je:

RePEc Short-ID: prb255

Záznam lze zobrazit v subsystému IDEAS (<http://ideas.repec.org/f/pbr255.html#details>) nebo také v subsystému NEP (<http://econpapers.repec.org/scripts/search.asp?ft=Brazdik>).

Druhá služba, známá pod zkratkou **EDIRC** (Economics Departments, Institutes and Research Centers in the World, <http://edirc.repec.org/>), zajišťuje správu autoritních informací o všech institucích (korporacích), ze kterých afilačně pocházejí autoři systému RePEc. V současnosti je v databázi této služby registrováno přibližně **11 000 institucí** z cca **229 zemí** (ekonomické ústavy nebo katedry univerzit, výzkumná centra, ministerstva financí, statistické úřady, banky aj.). Správcem této služby je Christian Zimmerman (Connecticutské univerzity, USA). Jednotlivých korporací je v autoritním záznamu přidělován také **jedinečný identifikátor**. Využívá se také v různých subsystémech a službách systému RePEc. Každý identifikátor korporace obsahuje celkem 7 znaků (<http://edirc.repec.org/handle.html>): **RePEc:edi:iiiiic**

Personal Details
First Name: Frantisek
Middle Name:
Last Name: Brazdik
Suffix:
RePEc Short-ID: pbr255
Email: [This author has chosen not to make the email address public]
Homepage: http://home.cerge-ei.cz/brazdik/
Postal Address: ČNB Na Příkopě 28 Praha 1 11503 Czech Republic
Phone: +420224414308
Affiliation
(in no particular order)
<ul style="list-style-type: none"> Center for Economic Research and Graduate Education (CERGE) Charles University Location: Praha, Czech Republic Homepage: http://www.cerge.cuni.cz/ Email: webmaster@cerge.cuni.cz Phone: (420-2) 240 05 111 Fax: (420-2) 242 27 143 Postal: Politických veznu 7, P.O. Box 882, 111 21 Praha 1 Handle: RePEc:edi:cergecz (registered authors at this institution) Center for Economic Research and Graduate Education and Economics Institute (CERGE-EI) Location: Praha, Czech Republic Homepage: http://www.cerge-ei.cz/ Email: library@cerge-ei.cz

Obr. č. 5/23: záznam českého doktoranda (CERGE-EI) s identifikátorem v databázi subsystému IDEAS digitálního systému RePEc (prb255) [převzato z WWW 2008-10-06]

Prvních pět znaků označuje instituci, další dva znaky reprezentují zemi či stát (užívá se kódů ISO pro internetové domény, například ekonomické asociace mají kód „ea“:

RePEc:edi:aeaaaaea (American Economic Association)

Příklady přidělených identifikátorů některých institucí z ČR (z rejstříku služby EDIRC, <http://edirc.repec.org/alphabet.html>):

RePEc:edi:uevsecz (Vysoká škola ekonomická, VŠE)

RePEc:edi:ffvsecz (VŠE, Fakulta Financí a Účetnictví, FFÚ)

RePEc:edi:dbvsecz (VŠE, FFÚ, Katedra bankovníctví a pojišťovnictví)

5.6.6 Závěr

Mohutný rozvoj informačních systémů všech typů, zejména pak systémů digitálních archivů, je v posledních letech doprovázen značným zintenzivněním snah po zavedení trvalé a globálně platné identifikace tvůrců, výrobců a distributorů informačních obsahů ve společnosti. Zajímavé technologické novinky řešení problémů jednoznačné identifikace autoritních entit předvádějí v poslední době knihovnické komunity, které novým způsobem těží z dat, jež po desetiletí ukládaly v bázích katalogizačních a autoritních záznamů. Výsledky jsou pozitivní, nicméně tato komunita je schopna identifikovat prozatím jen malou část všech tvůrců informačních obsahů. Řada odborníků proto v tuto chvíli vkládá velká očekávání do právě připravované mezinárodní normy ISO 27729 (ISPI), která by mohla značně napomoci

procesu jednoznačné a globálně platné identifikace autoritních entit v univerzálním záběru. Nicméně, mnozí již podnikli vlastní pokusy o zavedení moderní identifikace autorů. Inspirativní řešení jednoznačné a globálně platné identifikace autorů ve vědě předložila v tomto roce společnost Thomson Scientific (projekt ResearcherID). Jde však prozatím o dobrovolnou záležitost, perspektivy ale vypadají slibně. V oblasti digitálních archivů existuje celá řada dílčích řešení jednoznačného, avšak jen lokálně platného identifikování autorů a korporací. Zdařilým projektem v oblasti identifikace autorů v oblasti vědy a výzkumu poslední doby je rozhodně nizozemský národní systém DAI, který zavedl globálně platný a trvalý digitální identifikátor pro autory (s využitím nového URI „info“ schématu). Projekt mimořádně úspěšně spojil úsilí knihovníků a správců systémů pro registraci výsledků z vědy a výzkumu. Webová aplikace směřování identifikátorů DAI se právě připravuje a případné propojení s mezinárodním identifikátorem ISNI rovněž. Modelové řešení je natolik úspěšné, že se stalo vzorem pro budování jedné z komponent evropského systému DRIVER II. Pokud by tento evropský projekt týkal v budoucnu i ČR, bylo by rozumné přijmout již hotová řešení v oblasti trvalé identifikace autorů i u nás.

6 Závěry práce

Disertační práce *Digitální knihovny s volným přístupem v oblasti vědy a výzkumu a identifikace a metadatový popis jejich objektů* přinesla následující výsledky:

- V rovině teoretické stanovila pracovní definice a výklady nosných termínů ve zvoleném tématu, které nemají definitivní charakter, neboť jde o velmi dynamickou oblast a díky rozvoji nových technologií lze očekávat změny i v budoucnu
- Volně dostupné digitální knihovny nebo elektronické archivy jsou v současnosti budovány a rozvíjeny především sférou vysokoškolských a akademických pracovišť, a slouží prioritně komunitám jejich vlastních uživatelů. Zahnují typově dokumenty šedé literatury (preprinty, výzkumné a technické zprávy a disertace). Jak doložila analýza, jejich fondy jsou zatím ve srovnání s fondy digitálních knihoven komerční sféry malé, nicméně narůstají a některé z velkých digitálních knihoven již komerčním produktům konkurují či je také v některých aspektech předhánějí. Specifický význam mají nově zakládané citační služby nad digitálními knihovnami a elektronickými archivy, které mají zatím spíše experimentální ráz, nicméně jsou příslibem do budoucna. Předpokládá se, že mohou být v budoucnu nasazovány i do reálných hodnocení výsledků vědecké práce akademických a univerzitních pracovišť
- Specifickým praktickým přínosem práce je příprava a představení národních metadatových standardů pro popis a komunikaci vysokoškolských kvalifikačních prací a souvisejících entit. Jde zároveň o důležitý příspěvek k rozvoji českých systémů digitálních knihoven VŠKP a budování národního registru těchto prací. Autorka připojila rozvahu a diskusi k aplikaci trvalých identifikátorů všech zmiňovaných informačních objektů

Použité informační zdroje

- [AERTS, Raf](#). 2008a. Digital identifiers could keep up with authors' moves. *Nature : the international weekly journal of science*. 31 Jul 2008, vol. 454, no. 7204, s. 575-575. Dostupný také komerčně z WWW (DOI): <<http://dx.doi.org/10.1038/454575c>>. ISSN 0028-0836 (Print). ISSN 1476-4687 (Online).
- [AERTS, Raf](#). 2008b. Digital identifiers work for articles, so why not for authors? *Nature : the international weekly journal of science*. 19 Jun 2008, vol. 453, no. 7198, s. 979-979. Dostupný také komerčně z WWW (DOI): <<http://dx.doi.org/10.1038/453979b>>. ISSN 0028-0836 (Print). ISSN 1476-4687 (Online).
- ANSI/NISO Z39.84-2000. *Syntax for the Digital Object Identifier*. Bethesda (MD) : National Information Standards Organization Press, 2000. 11 s. Dostupný také z WWW: <<http://www.niso.org/standards/resources/Z39-84-2000.pdf>>. ISBN 1-880124-47-5. ISSN 1041-5653.
- ARMS, W.Y. 1995a. *Key Architectural Issues in The Digital Library* [online prezentace]. Reston (VA) : CNRI, February 1, 1995 [cit. 2008-12-12]. Dostupný z WWW: <<http://www.cnri.reston.va.us/home/cstr/arch/slides.html>>.
- ARMS, W.Y. 1995b. Key Concepts in the Architecture of the Digital Library. *D-Lib Magazine* [online]. July 1995 [cit. 2008-12-12]. Dostupný z WWW (HDL): <<http://hdl.handle.net/hdl:cnri.dlib/july95-arms>>. ISSN 1082-9873.
- ARMS, W.Y.; BLANCHI, Ch.; OVERLY, E.A. 1997. An Architecture for Information in Digital Libraries. *D-Lib Magazine* [online]. February 1997 [cit. 2008-12-12]. Dostupný z WWW (HDL): <<http://hdl.handle.net/hdl:cnri.dlib/february97-arms>>. ISSN 1082-9873.
- ARMS, W.Y. 2000a. Automated Digital Libraries : How Effectively Can Computers Be Used for the Skilled Tasks of professional Librarianship? *D-Lib Magazine* [online]. July/August 2000, vol. 6, no. 7/8 [cit. 2008-12-12]. Dostupný z WWW (DOI): <<http://dx.doi.org/10.1045/july2000-arms>>. ISSN 1082-9873.
- ARMS, W.Y. 2000b. *Digital Libraries*. Cambridge (Mass) : MIT Press, 2000. x, 287 s. Digital libraries and electronic publishing. Online vydání dostupné od roku 2005 také na WWW: <<http://www.cs.cornell.edu/wya/DigLib/>>. ISBN 0-262-01180-8. ISBN 978-0-262-01180-8.
- ASCHENBRENNER, Andreas aj. 2008. The Future of Repositories? : Patterns for (Cross-) Repository Architectures. *D-Lib Magazine* [online]. November/December, 2008, vol. 14, no. 11/12 [cit. 2008-12-12]. Dostupný z WWW (DOI): <<http://dx.doi.org/10.1045/november2008-aschenbrenner>>. ISSN 1082-9873.
- Asociace knihoven vysokých škol České republiky. 2004a. *Důvodová zpráva k otázkám budování systémů pro zpřístupňování elektronických vysokoškolských kvalifikačních prací v České republice* [online]. 2004-03-25 [cit. 2009-03-25]. 10 s. Dostupný z WWW: <<http://www.akvs.cz/pdf/vskpduvodovazprava.pdf>>.
- Asociace knihoven vysokých škol České republiky. 2004b. *K otázkám budování systémů pro zpřístupňování elektronických vysokoškolských kvalifikačních prací v České republice : Stanovisko Asociace knihoven vysokých škol* [online]. 2004 [cit. 2009-03-25]. 3 s. Dostupný z WWW: <<http://www.akvs.cz/pdf/vskpshrnuti.pdf>>.
- Association for Computing Machinery. 2003. *Computing Research Repository (CoRR)* [online]. Last updat. 2003-02-09 [cit. 2008-12-12]. Dostupný z WWW: <<http://www.acm.org/corr/>>.
- BARRUECO CRUZ, José Manuel; KLINK, Markus J. R.; [KRICHEL, Thomas](#). 2000a. *Personal data in a large digital library*. In *Research and Advanced Technology for Digital Libraries 4th European Conference, ECDL 2000 Lisbon, Portugal, September 18-20, 2000 : Proceedings*.

Berlin : Springer, 2000, s. 127. Lecture notes in computer science, 1923/2000. ISBN 978-3-540-41023-2. Dostupný také z WWW: <<http://openlib.org/home/krichel/phoenix.a4.pdf>>. Preprint dostupný také z WWW: <<http://eprints.rclis.org/435/>>. ISBN 978-3-540-41023-2.

- BARRUECO CRUZ, José Manuel; [KRICHEL, Thomas](#). 2000b. Cataloging Economics Preprints : An Introduction to the RePEC Project. *Journal of Internet Cataloguing*. 2000, vol. 3, issue 2/3, s. 227-241. Dostupný komerčně také z WWW: <<http://search.ebscohost.com/>>.
- BARRUECO CRUZ, José Manuel; SUBIRATS COLL, Imma. 2003a. RCLIS : towards a digital library for Information Science. In *Proceedings Libraries in Digital Age (LIDA), Dubrovnik and Mljet (Croatia)*. 2003-11-27. Preprint a prezentace dostupné také z E-LIS: <<http://eprints.rclis.org/352/>>.
- BARRUECO CRUZ, José Manuel; SUBIRATS COLL, Imma. 2003b. Documents in Information Science : a free space for our profession on the Internet. *Serials : The journal for the serials community*. 2003, vol. 16, issue 2, s. 211-214. Dostupný také z WWW: <<http://uksg.metapress.com/index/YYT7U83QBB17HVLA.pdf>>. ISSN 0953-0460.
- BARRUECO CRUZ, José Manuel; [KRICHEL, Thomas](#). 2005. Building an autonomous citation index for GL : RePEc, teh Economics working papers case. *The grey journal*. 2005, vol. 1, no. 2, s. 91-97. Dostupný komerčně také z WWW: <<http://search.ebscohost.com/>>. ISSN 1574-1796.
- BARTOŠEK, M. 2004. Digitální knihovny – teorie a praxe. *Národní knihovna* [online]. 2004, roč. 15, č. 4, s. 233-254 [cit. 2008-12-12]. Dostupný z WWW: <<http://knihovna.nkp.cz/pdf/0404/0404233.pdf>> nebo <<http://knihovna.nkp.cz/NKKR0404/0404233.html>>. Autorský postprint z 2005-01-26 dostupný z E-LIS: <<http://eprints.rclis.org/5061/>>. ISSN 1214-0678 (Online). ISSN 0862-7487 (Print).
- BERMÈS, Emmanuelle. 2006. Des identifiants pérennes pour les ressources numériques : l'expérience de la BnF [online]. Paris : BnF, 5 Mai 2006 [cit. 2009-03-25]. 9 s. (PDF). Les identifiants pérennes à la Bibliothèque nationale de France. Dostupný z WWW: <<http://bibnum.bnf.fr/identifiants/identifiants-200605.pdf>>.
- BORGMAN, Ch. L. 2003a. *From Gutenberg to the global information infrastructure : access to information in the networked world*. 1st paperback ed. Cambridge (Mass) : The MIT Press, 2003. xviii, 324 s. Digital libraries and electronic publishing. ISBN 0-262-52345-0 (brož.).
- BORGMAN, Christine L. 2003b. The invisible library : paradox of the global information infrastructure - challenges faced by libraries and proposed research designs. *Library trends*. May 2003, vol. 51, issue 4, s. 652-674. Dostupný také z WWW: <http://findarticles.com/p/articles/mi_m1387/is_4_51/ai_105046545>. ISSN 0024-2594.
- [BOURDON, Françoise](#). 1993. *International cooperation in the field of authority data : an analytical study with recommendations*. München : Saur, 1993. 135 s. UBCIM publications, New series, vol. 11. ISBN 3-598-11169-X.
- BRAND, Amy. 2004. CrossRef and the research experience. *Learned publishing*. 2004, vol. 17, no. 3, s. 225-230. Dostupný z WWW (DOI): <<http://dx.doi.org/10.1087/095315104323159658>>.
- BRAND, Amy; DALY, Frank; MEYERS, Barbara. 2003. *Metadata demystified : A Guide for Publishers*. Bethesda (MD) : NISO Press and The Sheridan Press, July 2003. 15 s. Dostupný volně také na WWW: <http://www.niso.org/publications/press/Metadata_Demystified.pdf>. ISBN 1-880124-59-9.
- BRANDEJS, Michal aj. 2008. Národní registr VŠKP a systém na odhalování plagiátů. In *INFORUM 2008 : 14. ročník konference o profesionálních informačních zdrojích , Praha 28.-30. května 2008 : sborník příspěvků z konference* [online]. Praha : Albertina icome Praha; Vysoká škola ekonomická v Praze, 2008 [cit. 2009-03-25]. [9 s.]. Dostupný z WWW: <<http://www.inforum.cz/pdf/2008/brandejs-michal-cze.pdf>>. ISSN 1801-2213.

- BRATKOVÁ, E. 1999. Metadata jako nový nástroj pro komunikaci webovských informačních zdrojů. *Národní knihovna* [online]. 1999, roč. 10, č. 4 [2008-12-12], s. 178-195. Dostupný z WWW: <<http://full.nkp.cz/nkkr/Nkkr9904/9904178.html>> nebo <<http://full.nkp.cz/nkkr/pdf/9904/9904178.pdf>>. ISSN 1214-0678 (Online). ISSN 0862-7487 (Print).
- BRATKOVÁ, E. 2000. Elektronické archívy vědecké literatury a jejich integrace. In *INFOS 2000 : zborník z 30. medzinárodného informatického sympózia , ktoré sa konalo v dňoch 3.-6. apríla 2000 v Starej Lesnej*. Bratislava : Spolok slovenských knihovníkov, 2000, s. 187-200. Dostupný také z WWW: <<http://www.aib.sk/infos/infos2000/19.htm>>. ISBN 80-81565-80-5.
- BRATKOVÁ, E. 2001. Citace odborné literatury jako nástroj rozvoje služeb a integrace digitálních knihoven. In *AKP 2001 : Automatizace knihovnických procesů - 8 : Sborník z 8. ročníku semináře pořádaného ve dnech 24.-25. dubna 2001 v Liberci*. Praha : ČVUT, Výpočetní a informační centrum 2001, s. 109-120. Dostupný také v elektronické formě z WWW: <<http://knihovny.cvut.cz/akp/clanky/12.pdf>>. ISBN 80-01-02-366-4.
- BRATKOVÁ, E. 2002a. Elektronické disertace a projekty a systémy jejich zpřístupňování. *Národní knihovna* [online]. 2002, roč. 13, č. 4 [2008-12-12], s. 268-287. Dostupný z WWW: <<http://full.nkp.cz/nkkr/Nkkr0204/0204268.html>>. ISSN 1214-0678 (Online). ISSN 0862-7487 (Print).
- BRATKOVÁ, Eva. 2002b. Informační entity, jejich znaky a vztahy : podle modelu FRBR. In *Informační studia a knihovnictví v elektronických textech I*. [CD-ROM]. Editori R. Papík, M. Souček, A. Stöcklová. Praha : Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví, 2001 [i.e. 2002]. 60 s. Zpracováno v rámci programu VISK-2 (MK ČR). Dostupný také z WWW (v části archiv): <<http://texty.jinonice.cuni.cz/>>.
- BRATKOVÁ, Eva. 2003a. Principy, aktéři, modely, postupy a technologie v systémech elektronických disertací. In *INFORUM 2003 : 9. konference o profesionálních informačních zdrojích , 27.-29.5. 2003 : sborník z konference informačních profesionálů* [online]. Praha : Albertina icome Praha; Vysoká škola ekonomická v Praze, 2003 [cit. 2009-03-25]. [9 s.]. Dostupný z WWW: <http://www.inforum.cz/inforum2003/prispevky/Bratkova_Eva.pdf>. ISSN 1801-2213.
- BRATKOVÁ, Eva. 2003b. Rozvoj otevřených archivů elektronických tisků a alternativních bibliografických služeb v oboru Informační věda. *Národní knihovna* [online]. 2003, roč. 14, č. 4 [cit. 2008-12-12], s. 254-269. Dostupný z WWW: <<http://full.nkp.cz/nkkr/pdf/0304/0304254.pdf>> nebo <<http://full.nkp.cz/nkkr/Nkkr0304/0304254.html>>. Autorský preprint z 2004-01-26 dostupný také z archivu E-LIS: <<http://eprints.rclis.org/745/>>. ISSN 1214-0678 (Online). ISSN 0862-7487 (Print).
- BRATKOVÁ, Eva. (zprac.). 2004a. *Systémy pro rozšiřování a zpřístupňování elektronických disertačních prací : analýza současného stavu v zahraničí*. Praha, 2003-10-31, poslední aktualizace 2004-02-22. 59 s. Zpráva pro grantovou úlohu FRVŠ č. 2045/2003 (okruh E2) „E-DI“ (Digitální knihovna diplomových a disertačních prací UK v Praze). Řešitelská pracoviště: Ústřední knihovna Univerzity Karlovy v Praze a Knihovna Centra pro ekonomický výzkum a doktorské studium Univerzity Karlovy v Praze. Dostupný také z WWW: <<http://www.evskp.cz/Dokumentyver/systemy-analyza.pdf>>.
- BRATKOVÁ, E. 2004b. Systémy pro zpřístupňování elektronických vysokoškolských kvalifikačních prací v České republice : od prvních lokálních systémů k platformě jejich dalšího společného rozvoje. In HALKO, Peter (ed.). *Elektronický zborník príspevkov z 11. medzinárodného seminára CASLIN 2004 (Čingov - Slovenský raj, 5.-9. september 2004)* [CD-ROM]. Prešov : Združenie KOLIN a Přešovská univerzita v Prešove, 2005. 1 disk CD-ROM. ISBN 80-8068-337-9. – Dostupný také z WWW: Dostupný z WWW: <<http://www.pulib.sk/elpub/CASLIN04/Bratkova.pdf>>.
- BRATKOVÁ, E. 2005a. K otázkám pojmu, třídění a typologie internetových a webovských informačních zdrojů. *Národní knihovna*. 1998, roč. 9, č. 5, s. 262-275. Aktualizovaný a doplněný

postprint z 2005-04-04 k dispozici z archivu E-LIS: <<http://eprints.rclis.org/3558/>>. ISSN 0862-7487 (Print).

- BRATKOVÁ, Eva. 2006a. Otevřený archiv elektronických tisků E-LIS zve také české autory. In *INFORUM 2006 : 12. ročník konference o profesionálních informačních zdrojích, Vysoká škola ekonomická, Praha, 23.-25. května 2006* [online]. Praha : Albertina icome Praha, 2006 [cit. 2009-03-25]. [12 s.]. Dostupný z WWW: <http://www.inforum.cz/inforum2006/pdf/Bratkova_Eva2.pdf>. Postprint dostupný také z archivu E-LIS <<http://eprints.rclis.org/14547/>>. ISSN 1801-2213.
- BRATKOVÁ, Eva. 2006b. Otevřený přístup, digitální knihovny a citační služby. In *INFORUM 2006 : 12. ročník konference o profesionálních informačních zdrojích, Vysoká škola ekonomická, Praha, 23.-25. května 2006* [online]. Praha : Albertina icome Praha, 2006 [cit. 2009-03-25]. [21 s.]. Dostupný z WWW: <http://www.inforum.cz/inforum2006/pdf/Bratkova_Eva.pdf>. ISSN 1801-2213.
- BRATKOVÁ, Eva; MACH, J. 2006c. Standardizace metadat pro národní registr EVŠKP. In MIKULECKÁ, J.; SEDLÁČEK, J. (ed.). *Digitální knihovny : provoz a budování : sborník konference EUNIS-CZ, Špindlerův Mlýn, 28. až 30. května 2006*. Vyd. 1. s přílohou CD-ROM. Hradec Králové : Gaudeamus, 2006, s. 5-12. Dostupný také z WWW: <http://www.eunis.cz/Krkonoze2006/sbornik/dokumenty/DIGITALNI_KNHOVNY.pdf>. Dostupný také z E-ELIS: <<http://eprints.rclis.org/6594/>>. ISBN 80-7041-200-3. – Prezentace: <<http://www.eunis.cz/Krkonoze2006/dokumenty/Bratkova.pps>>.
- BRATKOVÁ, Eva. 2007a. Nový online přístup k databázi SIGLE (projekt OpenSIGLE). *Ikaros : elektronický časopis o informační společnosti* [online]. 2007, roč. 11, č. 10 [2009-03-25]. [12 s.]. Dostupný z WWW: <<http://www.ikaros.cz/?q=node/4346>>. URN-NBN:cz-ik4346. ISSN 1212-5075.
- BRATKOVÁ, Eva. 2007b. *Otevřený přístup a digitální knihovny v oblasti vědy a výzkumu : (vybrané systémy)* [online]. Verze 1.0. Praha : Ústav informačních studií a knihovnictví FF UK v Praze, prosinec 2008 [cit. 2009-03-25]. 126 s. Elektronické studijní texty ÚISK. Zpracováno v rámci programu Veřejné informační služby knihoven, VISK-2 (MK ČR), Mimoškolní vzdělávání knihovníků, č. 12. Dostupný z WWW: <<http://texty.jinonice.cuni.cz/>>.
- BRATKOVÁ, Eva. 2007c. *Sít' trvalých identifikátorů informačních entit* [online]. Verze 1.0. Praha : Ústav informačních studií a knihovnictví FF UK v Praze, prosinec 2007 [cit. 2009-03-25]. 53 s. Formát PDF. Elektronické studijní texty ÚISK. Zpracováno v rámci programu Veřejné informační služby knihoven, VISK-2 (MK ČR), Mimoškolní vzdělávání knihovníků, č. 12. Dostupný z WWW: <<http://texty.jinonice.cuni.cz/>>.
- BRATKOVÁ, Eva. 2008a. Trvalá identifikace fyzických osob a korporací v prostředí digitálních archivů : aktuální směry řešení v zahraničních systémech. In *Systémy pro zpřístupňování VŠKP : zkušenosti, možnosti, nabídky, potřeby : 3. ročník semináře konaného 7. 10. 2008 na VUT v Brně* [online]. Praha : Odborná komise pro otázky elektronického zpřístupňování VŠKP AKVŠ ČR, 2008-12-22 [cit. 2009-03-25]. Připraveno v rámci Centralizovaného projektu „Národní registr VŠKP a úložiště závěrečných prací se službou na odhalování plagiátů“, MŠMT C1/2008. Dostupný z WWW: <<http://www.evskp.cz/Seminar3/seminar3-Bratkova2a.pdf>>. ISSN 1803-7003. – Prezentace PPT: <<http://www.evskp.cz/Seminar3/seminar3-Bratkova2.pdf>>.
- BRATKOVÁ, Eva; MACH, Jan. 2008b. Metadatový standard EVSKP-MS, v. 1.1 pro popis VŠKP a standardy související. In *Systémy pro zpřístupňování VŠKP : zkušenosti, možnosti, nabídky, potřeby : 3. ročník semináře konaného 7. 10. 2008 na VUT v Brně* [online]. Praha : Odborná komise pro otázky elektronického zpřístupňování VŠKP AKVŠ ČR, 2008-12-22 [cit. 2009-03-25]. Připraveno v rámci Centralizovaného projektu „Národní registr VŠKP a úložiště závěrečných prací se službou na odhalování plagiátů“, MŠMT C1/2008. Dostupný z WWW: <<http://www.evskp.cz/Seminar3/seminar3-BratkovaMach.pdf>>. ISSN 1803-7003. – Prezentace PPT: <<http://www.evskp.cz/Seminar3/seminar3-Bratkova1.pdf>>.

- BRODY, T. 2003. Citebase Search : Autonomous Citation Database for e-Print Archives. In *Conference on Worldwide Coherent Workforce, Satisfied Users - New Services For Scientific Information - 17-19 September 2003, Oldenburg, Germany*. 2003. Dostupný také z ECS: <<http://eprints.ecs.soton.ac.uk/10677/>>.
- BURGESS, J. 2006. *Intute Collection Development Framework and Policy* [online]. Version 1.4. Approved by the Intute Policy and Strategy Forum 12/10/06. 1 March 2006, last updated 18 September 2006. 23 s. Dostupný z WWW: <http://www.intute.ac.uk/Intute_cdfp.doc>.
- CANESSA, E; ZENNARO, M. (ed.). 2008. *Science Dissemination using Open Access : A compendium of selected literature on Open Access*. 1st ed. Trieste : ICTP - The Abdus Salam International Centre for Theoretical Physics, Science Dissemination Unit, July 2008. x, 196 s. Elektronická verze dostupná také na WWW: <<http://sdu.ictp.it/openaccess/book.html>>. ISBN 92-95003-40-3.
- CARPENTER, Leona (co-ordinat.) et al. 2003. *OAI for Beginners - The Open Archives Forum online tutorial* [online]. [S.l.] : Open Archives Forum, c2003- , last modif. 2003-10-14 [cit. 2008-12-28]. Dostupný z WWW: <<http://www.oaforum.org/tutorial/>>.
- CARR, Leslie; BRODY, Tim. 2007. Size Isn't Everything : Sustainable Repositories as Evidenced by Sustainable Deposit Profiles. *D-Lib Magazine* [online]. July/August 2007, vol. 13, no. 7/8 [cit. 2008-12-12]. Dostupný z WWW (DOI): <<http://dx.doi.org/10.1045/july2007-carr>>. ISSN 1082-9873.
- COLEMAN, A.; BRACKE, P. 2003a. *DLIST : Building An International Scholarly Communication Consortium for Library and Information Science* [online]. 2003 [cit. 2004-01-15]. Dostupný z archivu DLIST <<http://dlist.sir.arizona.edu/1785/>>.
- COLEMAN, A.; MALONE, Ch.; BRACKE, P. 2003b. *DLIST : Digital Library of Information Science & Information Science* [online prezentace]. 2003-03-18 [cit. 2008-12-12]. 19 snímků. Dostupný z WWW: <<http://www.sir.arizona.edu/faculty/coleman/uallibrary.ppt>>.
- COLEMAN, A; MALONE, Ch.; BRACKE, P. 2003c. *DLISTConnection : Information and Technology Literacy Service for NSDL* [online]. 21 October 2003 [cit. 2008-12-12]. Dostupný z archivu DLIST: <<http://dlist.sir.arizona.edu/271/>>.
- Consultative Committee for Space Data Systems. 2002-2007. *Reference Model for an Open Archival Information System (OAIS)* [online]. Washington (D.C.) : CCSDS, January 2002, corr. September 2007 [2008-12-12]. Přeruš. str. Recommendation for Space Data System Standards, CCSDS 650.0-B-1. Blue Book, Issue 1. [1, viii, 139 s.]. Dostupný z WWW: <<http://public.ccsds.org/publications/archive/650x0b1.pdf>>. Schváleno také jako ISO 14721:2003.
- CorpCZ : *Metadatový soubor pro popis korporací* [online]. Zpracovatelé Eva Bratková, Jan Mach. Verze 1.0. Praha : Odborná komise pro otázky elektronického zpřístupňování VŠKP AKVŠ ČR, 2008-05-05 [cit. 2009-03-25]. HTML soubor. Připraveno v rámci Centralizovaného projektu „Národní registr VŠKP a úložiště závěrečných prací se službou na odhalování plagiátů“, MŠMT C1/2008. Dostupný z WWW: <<http://www.evskp.cz/standardy/corpcz/1.0/>>.
- COYLE, Karen. 2006. Identifiers : unique, persistent, global. *The journal of academic librarianship*. July 2006, vol. 32, no. 4, s. 428-431. Dostupný komerčně také z WWW (DOI): <<http://dx.doi.org/10.1016/j.acalib.2006.04.004>>. ISSN 0099-1333.
- DANSKIN, Alan; DIXON, Anne; DOCHERTY, Michael; [HILL, Amanda](#); MOORE, Richard. 2008. *A review of the current landscape in relation to a proposed Name Authority Service for UK repositories of research outputs* [online]. Prepared for the JISC Names Project. September 2007 - June 2008 [2008-10-30]. 38 s. Dostupný z WWW: <<http://names.mimas.ac.uk/documents/LandscapeReport26Jun2008.pdf>>.
- De ROBBIO, A. 2003. E-LIS: un Open Archive in Library and Information Science. *Bibliotime : Rivista elettronica per le biblioteche* [online]. Marzo 2003, anno 6, no. 1 [cit. 2008-12-12].

Dostupný z WWW: <<http://didattica.spbo.unibo.it/bibliotime/num-vi-1/derobbio.htm>>. Dostupný také z E-LIS: <<http://eprints.rclis.org/201/>>. ISSN 1128-3564.

- DIJK, Elly; BAARS, Chris; HOGENAAR, Arjan aj. 2008. Accessing grey literature in an integrated environment of scientific research information. *The grey journal*. Summer 2008, vol. 4, no. 2, s. 57-66. ISSN 1574-1796.
- *DCMI Metadata Terms* [online]. Dublin : OCLC, 2008-01-14 [cit. 2008-10-19]. Dostupný z WWW: <<http://dublincore.org/documents/2008/01/14/dcmi-terms/>>.
- DIEPOLD, P. aj. 2001. Vom DFG-Projekt „Dissertationen Online“ zu DissOnline.de. *Bibliotheksdienst*. 2001, Jg. 35, Heft 3, s. 299-306. Dostupný také z WWW: <http://bibliotheksdienst.zlb.de/2001/01_03_04.pdf>. ISSN 0006-1972.
- DOBRATZ, S. aj. 2001. SGML/XML-based electronic theses and dissertations : existing projects and standards. *Internet and Higher Education*. 2001, vol. 4, issue 2, s. 93-104. Dostupný také komerčně z ScienceDirect (DOI): <[http://dx.doi.org/10.1016/S1096-7516\(01\)00052-5](http://dx.doi.org/10.1016/S1096-7516(01)00052-5)>. ISSN: 1096-7516.
- DOBRATZ, S.; SCHULZ, M. 2000a. Elektronisches Publizieren an Universitäten -aktuelle Trends und zwei Tagungen aus bibliothekarischer Sicht. I., Online-Dissertationen : Bericht vom Dritten Internationalen Symposium für elektronische Dissertationen. *Bibliotheksdienst*. 2000, Jg. 34, Heft 7/8, s. 1260-1264. Dostupný také z WWW: <http://bibliotheksdienst.zlb.de/2000/2000_07u08_Informationsvermittlung04.pdf>. ISSN 0006-1972.
- DOBRATZ, S.; SCHULZ, M. 2000b. Elektronisches Publizieren an Universitäten -aktuelle Trends und zwei Tagungen aus bibliothekarischer Sicht. II., XML-basierte elektronische Publikationen in Universitätsbibliotheken - eine Frage von Dokumenttypdefinitionen ? : Projekte, Ansätze und Ergebnisse eines NDLTD-Workshops zu Dokumenttypdefinitionen für Hochschulschriften. *Bibliotheksdienst*. 2000, Jg. 34, Heft 7/8, s. 1264-1274. Dostupný také z WWW: <http://bibliotheksdienst.zlb.de/2000/2000_07u08_Informationsvermittlung05.pdf>. ISSN 0006-1972.
- *Dublin Core Metadata Element Set, Version 1.1: Reference Description* [online]. Dublin : OCLC, 2004-12-20 [cit. 2009-03-25]. Dostupný z WWW: <<http://dublincore.org/documents/2004/12/20/dces/>>.
- *Dublin Core Metadata Element Set, Version 1.1* [online]. Dublin : OCLC, 2008-01-14 [cit. 2008-10-19]. Dostupný z WWW: <<http://dublincore.org/documents/2008/01/14/dces/>>.
- *Dublin Core Resource Types : Structuralist DRAFT : July 24, 1997* [online]. Maintained by Roy Tennant. Berkeley, University of California, Sunsite, last updated 1998-02-27. Dokument byl ale již není k dispozici na WWW.
- Dublin Core Metadata Initiative. 1998. *Type Element Working Draft* [online]. Simon Cox, Rebecca Guenther, Diann Rusch-Feja; DC Resource Type & Format Working Group. 1998-10-23 [cit. 2009-03-25]. Dostupný z WWW: <<http://dublincore.org/documents/1998/10/23/type-element/>>.
- Dublin Core Metadata Initiative. 1999. *List of Resouce Types* [online]. Editor(s) Rebecca Guenther; DC Type Working Group. 1999-08-05. Dokument již není k dispozici na WWW.
- Dublin Core Metadata Initiative. 2000. *DCMI Type Vocabulary* [online]. Dublin : OCLC, 2000-07-11 [cit. 2009-03-25]. Dostupný z WWW: <<http://dublincore.org/documents/2000/07/11/dcmi-type-vocabulary/>>.
- Dublin Core Metadata Initiative. 2003a. *DCMI Type Vocabulary* [online]. Dublin : OCLC, 2003-02-12 [cit. 2009-03-25]. Dostupný z WWW: <<http://dublincore.org/documents/2003/02/12/dcmi-type-vocabulary/>>.

- Dublin Core Metadata Initiative. 2003b. *DCMI Type Vocabulary* [online]. Dublin : OCLC, 2003-11-19 [cit. 2009-03-25]. Dostupný z WWW: <<http://dublincore.org/documents/2003/11/19/dcmi-type-vocabulary/>>.
- Dublin Core Metadata Initiative. 2003c. *Encoding Guidelines* [online]. Dublin: OCLC, 2003-12-09 [cit. 2009-03-25]. Dostupný z WWW: <<http://www.dublincore.org/resources/expressions/>>.
- Dublin Core Metadata Initiative. 2004. *DCMI Type Vocabulary* [online]. Dublin : OCLC, 2004-06-14 [cit. 2009-03-25]. Dostupný z WWW: <<http://dublincore.org/documents/2004/06/14/dcmi-type-vocabulary/>>.
- Dublin Core Metadata Initiative. 2009. *Dublin Core Metadata Initiative (DCMI) : [WebSite]* [online]. c1995-2009 [cit. 2009-03-25]. Dostupný z WWW: <<http://dublincore.org/>>.
- *ECS EPrints Repository* [online]. Southampton : University of Southampton, School of Electronics and Computer Science, c2008 [cit. 2008-12-12]. Dostupný z WWW: <<http://eprints.ecs.soton.ac.uk/>>.
- EDMINSTER, J.R. 2002a. *The Diffusion of New Media Scholarship : Power, Innovation, and Resistance in Academe* [online]. Tampa (Florida), May 2002 [cit. 2008-08-20]. Disertace (PhD). University of South Florida, Department of English. Dostupný přes NDLTD: <<http://purl.fcla.edu/fcla/etd/SFE0000035>>.
- EDMINSTER, J.R.; MOXLEY, J. 2002b. Graduate Education and the Evolving genre of Electronic Theses and Dissertations. *Computers and Composition*. 2002, vol. 19, s. 89-104. Dostupný také komerčně přes Science Direct: <[http://dx.doi.org/10.1016/S8755-4615\(02\)00082-8](http://dx.doi.org/10.1016/S8755-4615(02)00082-8)>. doi: 10.1016/S8755-4615(02)00082-8. ISSN: 8755-4615.
- *E-LIS : Eprints in Library and Information Science* [online]. 2003- [cit 2008-12-12]. V systému EPrints, verze 3.1.0. Dostupný z WWW: <<http://eprints.rclis.org/>>.
- *E-LIS Submission Guidelines* [online]. Version 2.0. Last modified November 23, 2005 [cit 2008-12-12]. Dostupný z WWW: <<http://elisdocs.rclis.org/guidelines/>>.
- EPrints.org. 2005. *EPrints.org* [online]. Southampton : University of Southampton, [2008] [cit. 2008-12-12]. Open Access. Self-Archiving FAQ. Dostupný z WWW: <<http://www.eprints.org/openaccess/self-faq/>>.
- ESPECHE, N. 2006. *ISI Web of Knowledge : A Commitment to Innovation* [online prezentace]. Helsinki (Finsko), May 2006 [2006-05-2]. Dokument byl dostupný z WWW: <http://www.lib.helsinki.fi/finelib/koulutus/ISI_Finland%20Presentation2006-final%5B1%5D.ppt>.
- *ETD-MS : an Interoperability Metadata Standard for Electronic Theses and Dissertations* [online]. Editors A. Atkins, E. Fox, R. France, H. Suleman. Version 1.00, revision 2. 2001-06-05, last updated on 2008-06-25 [cit. 2008-10-19]. Dostupný z WWW: <<http://www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.html>>.
- *EVSKP-MS : Metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR* [online]. Zpracovatelé Eva Bratková, Jan Mach. Verze 1.0. Praha : Odborná komise pro otázky elektronického zpřístupňování VŠKP AKVŠ ČR, 2008-05-05 [cit. 2009-03-25]. HTML soubor. Připraveno v rámci Centralizovaného projektu „Národní registr VŠKP a úložiště závěrečných prací se službou na odhalování plagiátů“, MŠMT C1/2008. Dostupný z WWW: <<http://www.evskp.cz/standardy/evskp/1.0/>>.
- *EVSKP-MS : Metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR* [online]. Zpracovatelé Eva Bratková, Jan Mach. Verze 1.1. Praha : Odborná komise pro otázky elektronického zpřístupňování VŠKP AKVŠ ČR, 2008-07-15 [cit. 2009-03-25]. HTML soubor. Připraveno v rámci Centralizovaného projektu „Národní registr VŠKP a úložiště závěrečných prací se službou na odhalování plagiátů“, MŠMT C1/2008. Dostupný z WWW: <<http://www.evskp.cz/standardy/evskp/1.1/>>.

- *EVSKP-MS : Metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR : Návrh* [online]. Zpracovatelé E. Bratková, E. Bulínová, J. Mach. Verze 0.1. Praha : Odborná komise pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací, Asociace knihoven vysokých škol ČR, 2005-11-07 [2009-03-25]. 23 s. Ve formátu HTML. Dostupný z WWW: <<http://www.evskp.cz/standardy/metadata/navrh-2005-11-07.html>>.
- FOX, E.A.; EATON, J.L.; McMILLAN, G. 1996. National Digital Library of Theses and Dissertations : A Scalable and Sustainable Approach to Unlock University Resources. *D-Lib Magazine* [online]. September 1996 [cit. 2008-08-20]. Dostupný z WWW: <<http://www.dlib.org/dlib/september96/theses/09fox.html>>. ISSN 1082-9873.
- FOX, E.A.; EATON, J.L.; McMILLAN, G. 1997. Networked Digital Library of Theses and Dissertations : An International Effort Unlocking University Resources. *D-Lib Magazine* [online]. September 1997 [cit. 2008-08-20]. Dostupný z WWW: <<http://www.dlib.org/dlib/september97/theses/09fox.html>>. ISSN 1082-9873.
- FRANCE, R.K. 2001. *Effective, Efficient Retrieval in a Network of Digital Objects* [online]. Blacksburg (Virginia), November 2001 [cit. 2008-09-20]. Disertace (PhD). Virginia Polytechnic Institute and State University. Dostupný přes NDLTD: <<http://scholar.lib.vt.edu/theses/available/etd-11272001-124212/unrestricted/RKFetd.pdf>>. URN etd-11272001-124212.
- GALLEZOT, G.; CHARTRON, G.; NOYER, J.-M. 2002. *Une archive ouverte des publications en InfoCom* [online]. 17 May 2002 [cit. 2008-12-12]. Preprint z kolokvia „Place et enjeux des revues pour la recherche en infoCom (SFSIC)“, 25 March 2002. Dostupný z @rchive SIC: <http://archivesic.ccsd.cnrs.fr/sic_00000025/fr/>.
- GALLEZOT, G. 2003. *ArchiveSIC, Archive Ouverte en Sciences de l'Information et de la Communication : rôle, fonctionnement et usage* [online]. 2003-09-09 [cit. 2008-12-12]. Publikováno v příručce Guide pratique : Les publications électroniques (Archimag, 2003). Dostupný z @rchive SIC: <http://archivesic.ccsd.cnrs.fr/sic_00000600/fr/>.
- GARTNER, Richard. 2008. *Metadata for digital libraries : state of the art and future directions* [online]. Version 1.0. Bristol : JICS, April 2008 [cit. 2009-03-25]. 19 s. Technology and standards watch. Dostupný z WWW: <http://www.jisc.ac.uk/media/documents/techwatch/tsw_0801pdf.pdf>.
- GREENBERG, D. 1998. Camel drivers and gatecrashers : quality control in the digital research library. In HAWKINS, B.L.; BATTIN, P. (ed.). *The mirage of continuity : reconfiguring academic information resources for the 21st century*. Washington (D.C.) : Council on Library and Information Resources; Association of American Universities, 1998, s. 105-116.
- *Guidelines for Digitization* [online]. Edited by Katherine M. Wisser. 2007 revised edition. Updated May 16, 2007 [cit. 2008-07-03]. North Carolina ECHO, Exploring Cultural Heritage Online. Glossary. Dostupný z WWW: <<http://www.ncecho.org/dig/digguidelines.shtml>>
- HAGEN, J.H.; DOBRATZ, S; SCHIRMBACHER, P. 2003. Electronic Theses and Dissertations Worldwide : Highlights of the ETD 2003 Symposium. *D-Lib Magazine* [online]. 2003, vol. 9, no. 7/8 [cit. 2008-08-20]. Dostupný z WWW: <<http://www.dlib.org/dlib/july03/hagen/07hagen.html>>. doi: 10.1045/july2003-hagen. ISSN 1082-9873.
- HAKALA, Juha. 2003. Popis dokumentů a přístup k nim - nové výzvy. In KLOUČKOVÁ, Z. a MACHALOVÁ, L. (sest.). *Moderní informační a komunikační technologie v knihovnictví 2003 : sborník příspěvků*. [Přel. M. Svoboda]. Praha : Státní technická knihovna, 2003, s. 22-30. ISBN 80-86504-09-3. -- Text originálu dostupný z WWW (CASLIN 2001): <<http://klement.nkp.cz/Caslin/caslin01/sbornik/hakala.html>>.
- HARDY, R.; OPPENHEIM, Ch.; BRODY, T. aj. 2005. *Open Access Citation Information* [online]. September 2005 [cit. 2009-03-25]. Final Report - Extended Version. JISC Scholarly

Communication Group. Ve formátu DOC. Dostupný z repozitáře ECS:
<<http://eprints.ecs.soton.ac.uk/11536/>>.

- HARNAD, S. 2001. The self-archiving initiative : Freeing the refereed research literature online. *Nature : the international weekly journal of science*. 26 April 2001, vol. 410, s. 1024-25. Dostupný také komerčně z WWW (DOI): <<http://dx.doi.org/10.1038/35074210>>. Preprint dostupný také z CogPrints: <<http://cogprints.ecs.soton.ac.uk/archive/00001642/01/nature4.htm>>. ISSN 0028-0836 (Print). ISSN 1476-4687 (Online).
- HARNAD, S.; BRODY, T. 2004. Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine* [online]. June 2004, vol. 10, no. 6 [cit. 2008-12-12]. Dostupný z WWW: <<http://dx.doi.org/10.1045/june2004-harnad>>. ISSN 1082-9873.
- HARNAD, S. 2006. Publish or Perish - Self-archive to Flourish : The green Route to Open Access [online]. *ERCIM News*. January 2006, no. 64 [cit. 2008-12-12]. Dostupný z WWW: <http://www.ercim.org/publication/Ercim_News/enw64/harnad.html>.
- HAWKINS, D.T.; LARSON, S.E.; CATON, B.Q. 2003. Information Science Abstracts : Tracking the Literature of Information Science. Part 2, A New Taxonomy for Information Science. *Journal of the American Society for Information Science and Technology*. 2003, vol. 54, no. 8, s. 771-781. Dostupný licenčně také z WWW: <<http://dx.doi.org/10.1002/asi.10275>>. ISSN 1532-2882.
- [HILL, Amanda](#). 2008. *What's in a Name? : Prototyping a name Authority Service for UK repositories* [online]. 2008 [2008-10-30]. Preprint. Příspěvek přednesený na konferenci ISKO (The 10th International Conference of International Society for Knowledge Organization : Culture and Identity in Knowledge Organization, 5-8 August, 2008, Montréal). Dostupný z WWW: <http://names.mimas.ac.uk/documents/Names_ISKO2008_paper.pdf>.
- HILSE, Hans-Werner; KOTHE, Jochen. 2006. *Implementing persistent identifiers : overview of concepts, guidelines and recommendations*. London : Consortium of European Research Libraries; Amsterdam : European Commission on Preservation and Access, 2006. 57 s. Dostupný také z WWW (URN): <<http://nbn-resolving.de/urn:nbn:de:gbv:7-isbn-90-6984-508-3-8>>. ISBN 90-6984-508-3.
- HIOM, Debra. 2006. Retrospective on the RDN. *Ariadne* [online]. April 2006, issue 47 [cit. 2008-12-12]. Dostupný z WWW: <<http://www.ariadne.ac.uk/issue47/hiom/>>. ISSN 1361-3200.
- HITCHCOCK, Steve; BRODY, Tim; HEY, Jessie M.N.; CARR, Leslie. 2007. Digital Preservation Service Provider Models for Institutional Repositories : Towards Distributed Services. *D-Lib Magazine* [online]. May/June 2007, vol. 13, no. 5/6 [cit. 2008-12-12]. Dostupný z WWW (DOI): <<http://dx.doi.org/10.1045/may2007-hitchcock>>. ISSN 1082-9873.
- International DOI Foundation. 2006. *The DOI Handbook* [online]. Version 4.4.1. Kidlington (Oxford, GB) : IDF, October 2006 [cit. 2009-03-25]. Dostupný z WWW (DOI): <<http://dx.doi.org/10.1000/182>>.
- Internet Assigned Numbers Authority. 2007. *MIME Media Types* [online]. c1999-2001, updated 2007-03-06 [cit. 2009-03-25]. Dostupný z WWW: <<http://www.iana.org/assignments/media-types/>>.
- ISO 14721:2003. *Space data and information transfer systems -- Open archival information system -- Reference model*. 1st ed. Geneve : ISO, 2003. 141 s.
- ISO 26324:200?. *Information and documentation – Digital Object Identifier (DOI) system*. 1st ed. Geneva : ISO, 200?. Připravuje se k vydání.
- *ISO Project 27729 : International Standard Name (Party) Identifier* [online]. Last update 2007-07-04 [cit. 2008-09-30]. Byl dostupný z WWW: <<http://www.collectionscanada.ca/iso/tc46sc9/wg6/index.html>>.

- *ISO Project 27729 : International Standard Party Number (ISPI)* [online]. Last update 2006-09-09 [cit. 2008-09-30]. Byl dostupný z WWW: <http://www.collectionscanada.ca/iso/tc46sc9/27729.htm>.
- JACSÓ, P. 2004a. CiteBaseSearch, Institute of Physics Archive, and Google's Index to Scholarly Archive. *Online*. September/October 2004, vol. 28, no. 5, s. 57-60. Dostupný také z: WWW: <http://www.jacso.info/PDFs/jacso-CiteBase-InsPhysics-GS-28-5.pdf>. ISSN 0146-5422.
- JACSÓ, Péter. 2004b. IDEAS, LogEc, EconLit. *Online*. May/June 2004, s. 57-60. Dostupný také komerčně WWW: <http://search.ebscohost.com/>. ISSN 0146-5422.
- JACSÓ, P. 2004c. *Web of Science Citation Indexes* [online]. August 2004 [cit. 2009-03-25]. Gale – Reference Reviews. Péter's Digital Reference Shelf. Dostupný z WWW: <http://www2.hawaii.edu/~jacso/gale.htm>.
- JACSÓ, P. 2005a. As we may search – Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current science*. November 2005, vol. 89, no. 9, s. 1537-1547. Dostupný také z WWW: <http://www.ias.ac.in/currsci/nov102005/1537.pdf>. ISSN 0011-3891.
- JACSÓ, P. 2005b. *CiteSeer* [online]. November 2005 [cit. 2009-03-25]. Gale – Reference Reviews. Péter's Digital Reference Shelf. Dostupný z WWW: <http://www.jacso.info/gale/citeseer/citeseer.htm>.
- KAHN, R. E.; WILENSKY, R. 1995. *A Framework for Distributed Digital Object Services* [online]. 1995-05-13 [cit. 2008-12-12]. Dostupný z WWW: <http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>.
- KAHN, R. E.; WILENSKY, R. 2006. A Framework for Distributed Digital Object Services. *International Journal on Digital Libraries*. 2006, vol. 6, no. 2, s. 115-123. Poprvé publikováno 1995-05-13. Dostupný také ze systému SpringerLink (DOI): <http://dx.doi.org/10.1007/s00799-005-0128-x>.
- KAPPENBERG, Bernd. 2008. European Scientists need a digital author identifier. In *EUROScience* [online]. August 5th 2008 [cit. 2008-10-30]. Dostupný z WWW: <http://www.euroscience.org/?p=30>.
- KIRRIEMUIR, John aj. 1998. Cross-Searching Subject Gateways. *D-Lib Magazine* [online]. January 1998 [cit. 2008-12-12]. Dostupný z WWW (HDL): <http://hdl.handle.net/hdl:cnri.dlib/january98-kirriemuir>. ISSN 1082-9873.
- KORB, Nikola; ROTH, Andreas. 2006. *EPICUR : Uniform Resource Name (URN) - Strategy of the German National Library* [online]. 2008-01-02 [cit. 2009-03-25]. Dostupný také z WWW (URN): <http://nbn-resolving.de/urn:nbn:de:1111-200606309>.
- KREJČÍŘ, V. 2006. Systémy pro tvorbu digitálních knihoven. In *INFORUM 2006 : 12. ročník konference o profesionálních informačních zdrojích, Vysoká škola ekonomická, Praha, 23.-25. května 2006* [online]. Praha : Albertina icome Praha, 2006. [cit. 2008-12-12]. Dostupný z WWW: http://www.inforum.cz/inforum2006/pdf/Krejcir_Vlastimil.pdf. ISSN 1801-2213.
- KRICHEL, Thomas. 2000. *Guildford Protocol* [online]. Version 2000-02-03 [cit. 2008-12-12]. Dostupný z WWW: http://openlib.org/acmes/root/docu/guilp_2000-02-03.html.
- KRICHEL, Thomas aj. 2007. *ReDIF* [online]. Version 1, Revision of 2007-06-01 [cit. 2008-12-12]. Dostupný z WWW: http://openlib.org/acmes/root/docu/redif_1.html.
- *KTD : Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online databáze]. Praha : Národní knihovna České republiky, 2003- [cit. 2008-12-12]. Dostupný z WWW: <http://sigma.nkp.cz/cze/ktid>.
- KUNZE, John A. (own.). 2007. *ARK (Archival Resource Key)* [online]. Last reviewed February 18, 2009 [cit. 2009-03-25]. Inside CDL. Digital Library Building Blocks. Best Practices and Standards. Dostupný z WWW: <http://www.cdlib.org/inside/diglib/ark/index.html>.

- KUNZE, John A. 2003. *Towards Electronic Persistence Using ARK Identifiers* [online]. July 2003 [cit. 2009-03-25]. 9 s. Inside CDL. Digital Library Building Blocks. Best Practices and Standards. Dostupný z WWW: <<http://www.cdlib.org/inside/diglib/ark/arkcdl.pdf>>.
- KUNZE, J.; RODGERS, R. P. C. 2008. *The ARK Persistent Identifier Scheme* [online]. May 2008 [cit. 2009-03-25]. 44 s. CDL Draft draft-kunze-ark-14.txt. Inside CDL. Digital Library Building Blocks. Best Practices and Standards. Dostupný z WWW: <<http://www.cdlib.org/inside/diglib/ark/arkspec.txt>>.
- LAGOZE, C. 2000. *The Cornell Digital Library Research Group : Architectures and Policies for Distributed Digital Libraries* [online]. 2000 [cit. 2008-12-12]. Materiál pro: DLW17, Tsukuba, Japan, February 2000. Dostupný z WWW: <<http://www.cs.cornell.edu/lagoze/papers/DLW17/cdlrg.htm>>.
- LAGOZE, C. aj. (ed.). 2008. *The Open Archives Initiative Protocol for Metadata Harvesting* [online]. Protocol Version 2.0 of 2002-06-14, Document version 2008-12-07T20:42:00Z. 2008 [cit. 2008-12-12]. Dostupný z World Wide Web: <<http://www.openarchives.org/OAI/openarchivesprotocol.html>>.
- LAVOIE, Brian F. 2004. *The Open Archival Information System Reference Model : Introductory Guide* [online]. January 2004 [cit. 2008-12-12]. 20 s. Technology Watch Series Report (Digital Preservation Coalition), 04-01. Dostupný z WWW: <http://www.dpconline.org/docs/lavoie_OAIS.pdf>.
- LAWRENCE, S.; BOLLACKER, K.; GILES, C.L. 1999a. Indexing and Retrieval of Scientific Literature. In *CIKM 99 : Eight International Conference on Information and Knowledge Management, Kansas City, Missouri, November 2-6, 1999*. 1999, s. 139-146. Dostupný také z CiteSeerX: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.39.9445>>.
- LAWRENCE, S.; GILES, C.L.; BOLLACKER, K. 1999b. Digital Libraries and Autonomous Citation Indexing. *IEE Computer*. 1999, vol. 32, no. 6, s. 67-71. Dostupný také z CiteSeerX: <<http://citeseer.ist.psu.edu/aci-computer/aci-computer99.html>>.
- LLORET, René. 2007a. *Minutes of the ISPI meeting, New York, 23rd January 2007* [online]. 2007-01-31 [2008-10-30]. [7] s. ISO/TC 46/SC 9/WG 6, N 3. Dostupný z WWW: <<http://www.collectionscanada.ca/iso/tc46sc9/wg6/wg6n3.pdf>>.
- LLORET, René. 2007b. *Minutes of the ISPI meeting, Santiago de Compostela, 8th May 2007* [online]. 2007-05-19 [2008-10-30]. [5] s. ISO/TC 46/SC 9/WG 6, N 5. Dostupný z WWW: <<http://www.collectionscanada.ca/iso/tc46sc9/wg6/wg6n5.pdf>>.
- MARTELLO, A. 2006. *Selection of Content for the Web Citation Index™ : Institutional Repositories and Subject-specific Archives* [online]. [2004, aktualiz. 2006] [cit. 2006-04-28]. Byl dostupný z WWW: <<http://scientific.thomson.com/free/essays/selectionofmaterial/wci-selection/>>.
- *METADISS : Format des Metadatenatzes für Online-Hochschulschriften der Deutschen Bibliothek und des Projekts "Dissertationen Online"* [online]. Version 1.5. Stand 2003-05-05 [cit. 2009-03-25]. Dostupný z WWW: <http://deposit.ddb.de/metadiss.htm#top_I>.
- *MetaPers : Format des Metadatenatzes für Online-Hochschulschriften der Deutschen Bibliothek und des Projekts "Dissertationen Online"* [online]. Version 1.1. Stand 2003-07-07 [cit. 2008-10-19]. Dostupný z WWW: <http://deposit.ddb.de/metadiss.htm#top_II>.
- Mezinárodní federace knihovnických asociací a institucí. IFLA Study Group on the Functional Requirements for Bibliographic Records. 1998. *Functional Requirements for Bibliographic Records : Final Report*. IFLA. München : Saur, 1998. viii, 136 s. UBCIM publications, New series, vol. 19. Dostupný také z WWW: <<http://www.ifla.org/VII/s13/frbr/frbr.pdf>>. ISBN 3-598-11382-X.
- Mezinárodní federace knihovnických asociací a institucí. IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). 2005. *Functional Requirements for Authority Records : A Conceptual Model* [online]. 2005-06-15 [cit. 2008-10-30]. 60 s. Draft.

Archivní kopie dostupná z WWW: <http://cidoc.ics.forth.gr/docs/frbr_oo/frbr_docs/Franar-Conceptual-M-Draft-e.pdf>.

- [Mezinárodní federace knihovnických asociací a institucí](#). IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). 2007. *Functional Requirements for Authority Data : A Conceptual Model* [online]. 2007-04-01 [cit. 2008-10-30]. 70 s. Draft. Dostupný z WWW: <<http://www.ifla.org/VII/d4/franar-conceptual-model-2ndreview.pdf>>.
- [Mezinárodní federace knihovnických asociací a institucí](#). IFLA Working Group on GARE Revision. 2001 *Guidelines for Authority Records and References*. 2nd ed. München : Saur, 2001. xii, 46 s. UBCIM publications, New series, vol. 23. Dostupný také na WWW: <<http://www.ifla.org/VII/s13/garr/garr.pdf>>. ISBN 3-598-11504-0.
- NEUHAUS, Ch. aj. 2006. The Depth and Breadth of Google Scholar : An Empirical Study. *Portal : libraries and the academy* [online]. April 2006, vol. 6, no. 2, s. 127-141 [cit. 2009-03-25]. Dostupný z WWW: <http://muse.jhu.edu/journals/portal_libraries_and_the_academy/toc/pla6.2.html> ISSN 1530-7131
- *OAI Repository Explorer* [online]. Explorer version - 1.46c : protocol version - 1.0/1.1/2.0 : December 2006. Cape town : Department of Computer Science, University of Cape Town, 2006 [cit. 2009-03-25]. Dostupný z WWW: <<http://re.cs.uct.ac.za/>>.
- *Open Access : Opportunities and Challenges : A Handbook*. European Commission, Directorate-General for Research. Luxembourg : Office for Official Publications of the European Communities, 2008. 139 s. Elektronická verze ve formátu PDF dostupná také z WWW (DOI): <<http://dx.doi.org/10.2777/93994>>. ISBN 978-92-79-06665-8.
- PASKIN, Norman. 2007. *Digital Object Identifier (DOI®) System*. Kidlington (Oxford, GB) : IDF, 2007 [cit. 2009-03-25]. 12 s. Text pro plánovanou publikaci „Encyclopedia of Library and Information science“ (nakladatele Taylor & Francis Group). Dostupný z WWW: <<http://www.doi.org/overview/070710-Overview.pdf>>.
- *PersCZ : Metadatový soubor pro popis fyzických osob* [online]. Zpracovatelé Eva Bratková, Jan Mach. Verze 1.0. Praha : Odborná komise pro otázky elektronického zpřístupňování VŠKP AKVŠ ČR, 2008-05-05 [cit. 2009-03-25]. HTML soubor. Připraveno v rámci Centralizovaného projektu „Národní registr VŠKP a úložiště závěrečných prací se službou na odhalování plagiátů“, MŠMT C1/2008. Dostupný z WWW: <<http://www.evskp.cz/standardy/perscz/1.0/>>.
- POWEL, J.; FOX, E.A. 1998. Multilingual Federated Searching Across Heterogeneous Collections. *D-Lib Magazine* [online]. September 1998 [cit. 2008-08-20]. Dostupný z WWW: <<http://www.dlib.org/dlib/september98/powell/09powell.html>>. ISSN 1082-9873.
- Publishers International Linking Association. 2007. *DOI Name Information and Guidelines* [online]. January 22, 2009 [cit. 2009-03-25]. 9 s. CrossRef. Info for Publisher. Dostupný z WWW: <<http://www.crossref.org/02publishers/doi-guidelines.pdf>>.
- *Reference Model for an Open Archival Information System (OAIS) : An Initiative of NASA in Cooperation with ISO and the Consultative Committee for Space Data Systems (CCSDS) - 1995-2002* [online]. 2003 [cit. 2008-12-12]. Dostupný z WWW: <<http://www.dlib.indiana.edu/education/brownbags/fall2003/metsbb/oais.htm>>.
- RFC 1630. 1994. *Universal Resource Identifiers in WWW : A Unifying Syntax for the Expressions of Names and Addresses of Objects on the Network as used in the World-Wide Web* [online]. T. Berners-Lee. June 1994 [cit. 2009-03-25]. 28 s. Dostupný z WWW: <<ftp://ftp.rfc-editor.org/in-notes/rfc1630.txt>>. ISSN 2070-1721.
- RFC 1737. *Functional Requirements for Uniform Resource Names* [online]. K. Sollins, L. Masinter. December 1994 [cit. 2009-03-25]. 7 s. Dostupný z WWW: <<ftp://ftp.rfc-editor.org/in-notes/rfc1737.txt>>. ISSN 2070-1721.

- RFC 2046. 1996. *Multipurpose Internet Mail Extensions (MIME). Part Two, Media Types* [online]. N. Freed, N. Borenstein. November 1996 [cit. 2009-03-25]. 44 s. RFC (Request for Comments) series. Dostupný z WWW: <<ftp://ftp.rfc-editor.org/in-notes/rfc2046.txt>>. ISSN 2070-1721.
- RFC 2077. 1997. *The Model Primary Content Type for Multipurpose Internet Mail Extensions* [online]. S. Nelson, C. Parks. January 1997 [cit. 2009-03-25]. 13 s. RFC (Request for Comments) series. Dostupný z WWW: <<ftp://ftp.rfc-editor.org/in-notes/rfc2077.txt>>. ISSN 2070-1721.
- RFC 2141. *URN Syntax* [online]. R. Moats. May 1997 [cit. 2009-03-25]. 8 s. Dostupný z WWW: <<ftp://ftp.rfc-editor.org/in-notes/rfc2141.txt>>. ISSN 2070-1721.
- RFC 2288. *Using Existing Bibliographic Identifiers as Uniform Resource Names* [online]. C. Lynch, C. Presto, R. Daniel. February 1988 [cit. 2009-03-25]. 10 s. Dostupný z WWW: <<ftp://ftp.rfc-editor.org/in-notes/rfc2288.txt>>. ISSN 2070-1721.
- RFC 2396. 1998. *Uniform Resource Identifiers (URI) : Generic Syntax* [online]. T. Berners-Lee, R. Fielding, L. Masinter. August 1998 [cit. 2009-03-25]. 40 s. Dostupný z WWW: <<ftp://ftp.rfc-editor.org/in-notes/rfc2396.txt>>. ISSN 2070-1721.
- RFC 2611. *URN Namespace Definition Mechanisms* [online]. L. Daigle, D. van Gulik, R. Iannella, P. Falstrom. June 1999 [cit. 2009-03-25]. 14 s. Dostupný z WWW: <<ftp://ftp.rfc-editor.org/in-notes/rfc2611.txt>>. ISSN 2070-1721.
- RFC 3188. *Using National Bibliography Numbers as Uniform Resource Names* [online]. J. Hakala. October 2001 [cit. 2009-03-25]. 13 s. Dostupný z WWW: <<ftp://ftp.rfc-editor.org/in-notes/rfc3188.txt>>. ISSN 2070-1721.
- RFC 3986. *Uniform Resource Identifier (URI) : Generic Syntax* [online]. T. Berners-Lee, R. Fielding, L. Masinter. January 2005 [cit. 2009-03-25]. 61 s. Dostupný z WWW: <<ftp://ftp.rfc-editor.org/in-notes/rfc3986.txt>>. ISSN 2070-1721.
- RFC 4452. 2006. *The "info" URI Scheme for Information Assets with Identifiers in Public Namespaces* [online]. [H. Van de Sompel](#), T. Hammond, E. Neylon, [S. Weibel](#). April 2006 [cit. 2009-03-25]. 17 s. RFC (Request for Comments) series. Dostupný z WWW: <<ftp://ftp.rfc-editor.org/in-notes/rfc4452.txt>>. ISSN 2070-1721.
- RUSBRIDGE, Ch. 1998. Towards the Hybrid Library. *D-Lib Magazine* [online]. July/August 1998 [cit. 2008-12-12]. Dostupný z WWW (HDL): <<http://hdl.handle.net/hdl:cnri.dlib/july98-rusbridge>>. ISSN 1082-9873.
- ŘEPIŠOVÁ, Z.; VOCHOZKOVÁ, H. 2000. Elektronické disertace on-line. In *RUFIS 2000 : sborník z česko-slovenské konference pořádané ve dnech 5.-6. 9. 2000 v Brně*. Brno : Vysoké učení technické v Brně; Masarykova univerzita Brno, 2000, s. 105-110.
- SAWYER, Don; REICH, Lou. 2002. *ISO "Reference Model For an Open Archival Information System (OAIS)" : Tutorial Presentation* [online prezentace]. October 2002 [cit. 2008-12-12]. 61 snímků (ve formátu PPT). Dostupný z WWW: <http://nssdc.gsfc.nasa.gov/nost/isoas/presentations/oais_tutorial_200210.ppt>.
- SHAFER, Keith aj. 1996. Introduction to Persistent Uniform Resource Locators. In *INET96 : Montreal, Canada, 24-28 June 1996 : Transforming Our Society Now* [online]. 1996 [cit. 2009-03-25]. Dostupný z WWW: <http://www.isoc.org/inet96/proceedings/a4/a4_1.htm>. Dostupný také z WWW: <<http://purl.oclc.org/docs/inet96.html>>.
- [SNYMAN, M. M. M.](#); Van RENSBURG, M. Jansen. 2000. Revolutionizing name authority control. In *ACM 2000 digital libraries : proceedings of the Fifth ACM Conference on Digital Libraries, June 2-7, 2000, San Antonio, Texas*. Edited by Ken Anderson. New York : Association for Computing Machinery, 2000, s. 185-194. Dostupný komerčně také z WWW (DOI): <<http://doi.acm.org/10.1145/336597.336660>>. ISBN 1-58113-231-X.

- [SPANJE, Daniël van](#). 2007. *Digital Author Identification (UKSG 17 - 18 april 2007)* [online]. Groningen : Rijksuniversiteit Groningen, Bibliotheek, 4 May 2007 [2008-10-30]. Textový popis k prezentaci PPT představené na konferenci UKSG 2007 (United Kingdom Serial Group, 30th UKSG Annual Conference and Exhibition, 16-18 April 2007, University of Warwick, Workshops and Briefing Sessions). Dostupný z WWW: <http://dai-uitrol.ub.rug.nl/logboek/FILES/61/DAI_UKSG_20070417_final.doc>.
- SUBER, P. 2007. *Open Access Overview : Focusing on open access to peer-reviewed research articles and their preprints* [online]. First put online, June 21, 2004, Last revised June 19, 2007 [cit. 2008-12-12]. Dostupný z WWW: <<http://www.earlham.edu/~peters/fos/overview.htm>>.
- SUBER, P. 2008a. Open Access in the United States. In *Open Access : Opportunities and Challenges : A Handbook*. European Commission, Directorate-General for Research. Luxembourg : Office for Official Publications of the European Communities, 2008, s. 124-127. Elektronická verze ve formátu PDF dostupná také z WWW (DOI): <<http://dx.doi.org/10.2777/93994>>. ISBN 978-92-79-06665-8.
- SUBER, P. 2008b. *Timeline of the Open Access Movement* [online]. Last revised November 7, 2008 [cit. 2008-12-12]. Dříve pod názvem „Formerly called the Timeline of the Free Online Scholarship Movement“. Dostupný z WWW: <<http://www.earlham.edu/~peters/fos/timeline.htm>>.
- SULEMAN, H. aj. 2001a. Networked Digital Library of Theses and Dissertations : Bridging the Gaps for Global Access. Part 1, Mission and Progress. *D-Lib Magazine* [online]. 2001, vol. 7, no. 9 [cit. 2008-08-20]. Dostupný z WWW: <<http://www.dlib.org/dlib/september01/suleman/09suleman-pt1.html>>. doi: 10.1045/september2001-suleman-pt1. ISSN 1082-9873.
- SULEMAN, H. aj. 2001b. Networked Digital Library of Theses and Dissertations : Bridging the Gaps for Global Access. Part 2, Services and Research. *D-Lib Magazine* [online]. 2001, vol. 7, no. 9 [cit. 2008-08-20]. Dostupný z WWW: <<http://www.dlib.org/dlib/september01/suleman/09suleman-pt2.html>>. 10.1045/september2001-suleman-pt2. ISSN 1082-9873.
- SULEMAN, H.; FOX, E.A. 2001c. A Framework for Building Open Digital Libraries. *D-Lib Magazine* [online]. 2001, vol. 7, no. 12 [cit. 2008-08-20]. Dostupný z WWW: <<http://www.dlib.org/dlib/december01/suleman/12suleman.html>>. doi: 10.1045/december2001-suleman. ISSN 1082-9873.
- SULEMAN, H. 2002a. *Open Digital Libraries* [online]. Blacksburg (Virginia), November 19, 2002 [cit. 2008-08-20]. Disertace (PhD). Virginia Polytechnic Institute and State University. Dostupný přes NDLD: <<http://scholar.lib.vt.edu/theses/available/etd-11222002-155624/unrestricted/odl.pdf>>. URN etd-11222002-155624.
- SULEMAN, H.; FOX, E.A. 2002b. *Beyond Harvesting : Digital Library Components as OAI Extensions* [online]. January 2002 [cit. 2008-08-20]. Technical Report, TR-02-25. Virginia Tech Dept. of Computer Science. Dostupný z WWW: <http://eprints.cs.vt.edu/archive/00000625/01/cstr_2002_odl_1.pdf>.
- *TEF 2.0 : Les métadonnées des thèses électroniques françaises* [online]. Version 2.0. Recommandation préparée sous l'égide du groupe CG46/CN357/GE5 de l'AFNOR. Mars 2006 [cit. 2008-10-19]. Dostupný z WWW: <<http://www.abes.fr/abes/documents/tef/recommandation/index.html>>. Webové sídlo standardu dostupné z WWW: <<http://www.abes.fr/abes/documents/tef/index.html>>.
- *Terms of Reference for ISO TC 46/SC 9 Working Group 6 : ISO Project 27729, "International Standard Name Identifier (ISNI)"* [online]. 2006-09-09, revised 2007-10-20 [cit. 2008-10-30]. ISO/TC 46/SC 9/WG 6 N 1. Dostupný z WWW: <<http://www.collectionscanada.gc.ca/iso/tc46sc9/wg6/sc9wg6n1.pdf>>.

- [Thomson Scientific \(firma\)](#). 2008. *ResearcherID.com* [online]. c2008 [cit. 2008-10-30]. What is ResearcherID? Dostupný z WWW: <<http://www.researcherid.com/resources/html/faq.html>>.
- [TILLET, Barbara B.](#) 2000. Authority Control on the Web. In *Conference on Bibliographic Control in the New Millenium, Washington 2000* [online]. Washington (DC) : Library of Congress, 2000 [cit. 2008-10-30]. Dostupný z WWW: <http://lcweb.loc.gov/catdir/bibcontrol/tillett_paper.html>.
- [TILLET, Barbara B.](#) 2007. Numbers to Identify Entities (ISADNs - International Standard Authority Data Numbers). *Cataloguing and Classification Quarterly*. 2007, vol 44, no. 3/4, s. 343-361. Dostupný komerčně také z WWW (DOI): <http://dx.doi.org/10.1300/J104v44n03_12>. ISSN 0163-9374.
- [TILLET, Barbara B.](#) 2008. *A review of the Feasibility of an International Standard Authority Data Numbers* [online]. Prepared for the IFLA Working Group on Functional Requirements and Numbering of Authority Records; Edited by Glenn E. Patton. 1 Juli 2008, approved by Standing Committee of the IFLA Cataloguig Section, 15 September 2008 [cit. 2008-10-30]. Zpráva. Dostupná z WWW: <<http://www.ifla.org/VII/d4/franar-numbering-paper.pdf>>.
- *Understanding Metadata*. Bethesda (MD) : NISO Press, c2004. 16 s. Dostupný volně také na WWW: <<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>. ISBN 1-880124-62-9. ISBN 978-1-880124-62-8.
- Uniform Resource Names : A progress report. The URN Implementors. *D-Lib Magazine* [online]. February 1996 [cit. 2008-12-12]. Dostupný z WWW (HDL): <http://hdl.handle.net/cnri.dlib/february96-urn_implementors>. ISSN 1082-9873.
- *URI Resource Pages* [online]. Washington (D.C.) : Library of Congress, October 28, 2008 [cit. 2009-03-25]. IFLA/CDNL Aliance for Bibliographic Standards. Dostupný z WWW: <<http://www.loc.gov/standards/uri/>>.
- Van de SOMPEL, Herbert. 2000. The UPS Prototype : An Experimental End-User Service across E-Print Archives. *D-Lib Magazine* [online]. February 2000 [cit. 2008-12-12]. Dostupný z WWW (DOI): <<http://dx.doi.org/10.1045/february2000-vandesompel-ups>>. ISSN 1082-9873.
- WATERS, Donald J. 1998. What are digital libraries? *CLIR Issues* [online]. July/August 1998, no. 4 [cit. 2008-12-12]. Dostupný z WWW: <<http://www.clir.org/pubs/issues/issues04.html#dlf>>.
- [WILLER, Mirna](#). 1996. Authority control and International Standard Authority Data Numbers : need for international cooperation. In *Authority control in the 21st century : an invitational conference, March 31 - April 1, 1996 : proceedings* [online]. Dublin, OH : OCLC, 1996 [cit. 2008-10-30]. Dostupný také z digitálního archivu OCLC: <<http://digitalarchive.oclc.org/da/ViewObjectMain.jsp?fileid=0000003520:000000091799&reqid=354>>.
- WILLIAMS, Caroline. 2006. Intute : The New Best of the Web. *Ariadne* [online]. July 2006, issue 48 [cit. 2008-12-12]. Dostupný z WWW: <<http://www.ariadne.ac.uk/issue48/williams/>>. ISSN 1361-3200.
- WITTEN, Ian H.; CUNNINGHAM, Sally Jo; APPERLEY, Mark D. 1996. The New Zealand Digital Library Project. *D-Lib Magazine* [online]. November 1996 [cit. 2008-12-12]. Dostupný z WWW (HDL): <<http://hdl.handle.net/hdl:cnri.dlib/november96-witten>>. ISSN 1082-9873.
- WITTEN, Ian H.; BAINBRIDGE, David. 2000. *How to build a Digital Library*. Amsterdam ; San Francisco (CA) : Morgan Kaufman, 2000. xxxiii, 518 s. ISBN 1-55860-790-0.
- World Wide Web Consortium. *RDF/XML Syntax Specification (Revised) : W3C Recommendation 10 February 2004* [online]. 2004 [cit. 2006-05-05]. Dostupný z WWW: <<http://www.w3.org/TR/rdf-syntax-grammar/>>.
- [XIA, Jingfeng](#). 2006. Personal name identification in the practices of digital repositories. *Program : Electronic library and information systems*. 2006, vol. 40, issue 3, s. 256-267. Dostupný

komerčně také z WWW (DOI): <<http://dx.doi.org/10.1108/00330330610681330>>. ISSN 0033-0337. Postprint volně dostupný také z archivu DLIST: <<http://dlist.sir.arizona.edu/1832/>>.

- *XMetaDiss : Format des Metadatenatzes der Deutschen Nationalbibliothek für Online-Hochschulschriften inklusive Angaben zum Autor (XmetaPers)* [online]. Herausgegeben von der Koordinierungsstelle DissOnline. Version 1.3. Stand 2006-08-22 [cit. 2008-10-19]. Německá verze dostupná z WWW: <http://www.d-nb.de/standards/pdf/ref_xmetadiss_v1-3.pdf>. Anglická verze dostupná z WWW: <http://www.d-nb.de/eng/standards/pdf/ref_xmetadiss_v1-3.pdf>. Webové sídlo standardu: <<http://www.d-nb.de/standards/xmetadiss/xmetadiss.htm>>.
- ZUBER, Peter A. 2008. A Study of Institutional Repository Holdings by Academic Discipline. *D-Lib Magazine* [online]. November/December, 2008, vol. 14, no. 11/12 [cit. 2008-12-12]. Dostupný z WWW (DOI): <<http://dx.doi.org/10.1045/november2008-zuber>>. ISSN 1082-9873.

Seznam použitých zkratk

ACI	Autonomous Citation Indexing
ACM	Association for Computing Machinery
ADAM	Art, Design, Architecture & Media Information Gateway
ADK	Automatizovaná Digitální Knihovna
ADT	Australian Digital Theses a Australasian Digital Theses
AHRC	Arts and Humanities Research Council
AIP	Archival Information Package
AKVŠ	Asociace knihoven vysokých škol České republiky
ANR	Access to Network Resources
ARC	Cross Archive Search Service
ARK	Archival Resource Key
ARPA	Advanced Research Projects Agency
ARPANET	Advanced Research Projects Agency Network
ASCII	American Standard Code for Information Interchange
AUT	Authority (Databáze národních autorit NK ČR)
AWPE	Abstracts of Working Papers in Economics
BIZ/ED	Business Education on the Internet
BOAI	Budapest Open Access Initiative
CCSDS	Consultative Committee for Space Data Systems,
CDLRG	Cornell Digital Library Research Group
CERN	Conseil Européen pour la recherche nucléaire
CGI	Common Gateway Interfac
CiTEc	Citations in Economics
CNRI	Corporation for National Research Initiatives,
CoRR	Computing Research Repository
CogPrint	Cognitive Sciences Eprint Archive
CSTC	Computer Science Teaching Center
CSTR	Computer Science Technical Reports Project
DARPA	Defence Advanced Research Projects Agency
DBC	Databázová centra
DC	Dublin Core
DCMI	Dublin Core Metadata Initiative
DFG	Deutsche Forschungsgemeinschaft
DIAS	Dissemination Information Package
DIP	Digital Information Archiving System

DL	Digital Library
DLIST	Digital Library of Information Science and Technology
DK	Digitální Knihovna
DOAJ	Directory of Open Access Journals
DOI	Digital Object Identifier
DOIS	Documents in Information Science
DSEP	Deposit System for Electronic Publications
DVK	Digitální Virtuální Knihovna
E-Lib	Electronic Libraries (Programme)
E-LIS	Eprints in Library and Information Science
ECS	School of Electronics & Computer Science, University of Southampton
EDD	Electronic Document Delivery
EDIRC	Economics Departments, Institutes and Research Centers in the World
EECS	Edinburg Engineering Virtual Library
EEVL	Department of Electrical Engineering and Computer Science (Berkeley)
ENSSIB	École Nationale Supérieure des Sciences de l'Information et Bibliothéconomie
ERIC	Education Resources Information Center
ESRC	Economic and Social Research Council
EVŠKP	Elektronické vysokoškolské kvalifikační práce
FGDC	Federal Geographic Data Committee (USA)
FTP	File Transfer Protocol
GII	Global Information Infrastructure
HAL	Hyper Article en Ligne
HCI	Human-Computer Interaction
HDL	Handle (trvalý identifikátor)
HTML	HyperText Markup Language
HTTP	Internet Anonymous FTP Archive Working Group
IAFA	Hypertext Transfer Protocol
IANA	Internet Engineering Task Force
IETF	Internet Assigned Numbers Authority
JISC	Joint Information Systems Committee (Velká Británie)
IDEAS	Internet Documents in Economics Access Service (RePEc)
IFLA	International Federation of Library Associations and Institutions
IETF	Internet Engineering Task Force
IR	Institutional Repository (Institucionální Repozitář)
ISBN	International Standard Book Number
ISSN	International Standard Serial Number
ISO	International Organization for Standardization
JEL	Journal of Economic Literature
JPEG	Joint Photographic Experts Group (formát JPEG file format)

JSP	Java Server Pages
KTD	Česká terminologická databáze knihovnictví a informační vědy (TDKIV)
LISA	Library and Information Science Abstracts (CSA)
LISTA	Library, Information Science and Technology Abstracts (EBSCO)
MARC	MAchine-Readable Cataloging
METS	Metadata Encoding and Transmission Standard
MG	Managing Gigabytes
MIMAS	Manchester Information & Associated Services
MIME	Multipurpose Internet Mail Extensions
MIT	Massachusetts Institute of Technology
MPRA	Munich Personal RePEc Archive
MVS	Meziknihovní výpůjční služba
NACA	National Advisory Committee for Aeronautics Technical Report Server
NARA	National Archives and Records Administration (USA)
NASA	National Aeronautic and Space Administration
NDLTD	Networked Digital Library of Theses and Dissertations
NEP	New Economics Papers
NSCTRL	Networked Computer Science Technical Reference Library
NEDLIB	Národní knihovna České republiky
NK ČR	Networked European Deposit Library
NSF	National Science Foundation
NSSDC	National Space Science Data Center
NZDL	New Zealand Digital Library
OA	Open Access
OAI	Open Archives Initiative
OAI-PMH	Open Archives Initiative - Protocol for Metadata Harvesting
OAIS	Open Archival Information System
OCLC	Online Computer Library Center
OMNI	Organising Medical Networked Information
OpCit	Open Citation Project
OpenDOAR	Open - Directory of Open Access Repositories
PDF	Portable Document Format
PDI	Preservation Description Information
PREMIS	Preservation Metadata: Implementation Strategies
PURL	Persistent URL
RAP	Repository Access Protocol
RCLIS	Research in Computing, Library and Information Science
RDA	Resource Description and Access
RDF	Resource Description Framework
RDN	Resource Discovery Network

RDNC	Resource Discovery Network Centre
RE	Repository Explorer
RePEc	Research Papers in Economics
RFC	Reference Model for an Open Archival Information System
RLG	Research Libraries Group
RM OAIS	Request For Comments
ROADS	Resource Organization And Discovery in Subject-based services
ROAR	Registry of Open Access Repositories
RSS	Really Simple Syndication nebo RDF Site Summary či Rich Site Summary
RUDI	Resource for Urban Design Information
SCI	Science Citation Index
SDI	Selective Dissemination of Information
SGML	Standard Generalized Markup Language
SIC	Sciences de l'Information et de la Communication
SIP	Submission Information Package
SOSIG	Social Science Information Gateway
SPARC	Scholarly Publishing and Academic Resources Coalition
TDKIV	Česká terminologická databáze knihovnictví a informační vědy (KTD)
TEI	Text Encoding Initiative,
TIFF	Tag Image File Format
UMI	University Microfilms International
UML	Unified Modeling Language
URC	Uniform Resource Characteristic
URL	Uniform Resource Locator
URN	Uniform Resource Name
VŠKP	Vysokoškolské kvalifikační práce
WoPEc	Working Papers in Economics
WOK	Web of Knowledge
WOS	Web of Science
XML	eXtensible Markup Language

Přílohy

Příloha 1:

EVSKP-MS : Metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR.
Verze 1.1. 2008-07-15. 31 s.

Příloha 2

PersCZ : Metadatový soubor pro popis fyzických osob. Verze 1.0. 2008-05-05. 15 s.

Příloha 3

CorpCZ : Metadatový soubor pro popis korporací. Verze 1.0. 2008-05-05. 12 s.

Příloha 4

EVSKP-MS : Metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR.
Verze 0.1 (vývojová). 2005-11-07. 14 s.

EVSKP-MS

Metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR

Verze 1.1

Tato verze:

<http://www.evskp.cz/standardy/evskp/1.1/>

Aktuální verze:

<http://www.evskp.cz/standardy/evskp/>

Předcházející verze:

<http://www.evskp.cz/standardy/evskp/1.0/>

Zpracovatelé:

Eva Bratková (Eva.Bratkova@ff.cuni.cz)

Jan Mach (machj@vse.cz)

Praha

15. července 2008

Obsah

Citace	3
Abstrakt	3
Status dokumentu	3
1 Úvod	4
1.1 Struktura popisu metadatových prvků	4
1.2 Jmenné prostory a XML schémata	5
1.3 Příklady zápisu	6
2 Definice prvků	7
2.1 Popisné metadatové prvky	7
evskp:metadata	7
dc:title	8
dcterms:alternative	9
dc:creator	10
dc:subject	11
dcterms:abstract	12
dcterms:tableOfContents	13
dc:publisher	13
dc:contributor	14
dcterms:created	15
dcterms:dateSubmitted	16
dcterms:dateAccepted	16
dcterms:modified	17
dc:type	17
dcterms:medium	18
dcterms:extent	19
dc:identifier	20
dc:language	21
dcterms:bibliographicCitation	21
dc:rights	22
thesis:degree	22
thesis:name (prvek vnořený v prvku thesis:degree)	23
thesis:level (prvek vnořený v prvku thesis:degree)	23
thesis:discipline (prvek vnořený v prvku thesis:degree)	24
thesis:grantor (prvek vnořený v prvku thesis:degree)	25
2.2 Technické a administrativní metadatové prvky	25
evskp:contact	26
evskp:fileNumber	26
evskp:fileProperties	27
evskp:transfer	28
evskp:server	29
evskp:dateDelivered	30
dcterms:available	30
evskp:modified	31

Citace

EVSKP-MS : metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR [online]. Zpracovatelé Eva Bratková, Jan Mach. Verze 1.1. Praha: Odborná komise pro otázky elektronického zpřístupňování VŠKP AKVŠ ČR, 2008-07-15 [2008-07-15]. Dostupný z WWW: <<http://www.evskp.cz/standardy/evskp/1.1/>>.

Abstrakt

Dokument definuje formální standardní soubor metadatových prvků určených pro popis vysokoškolských kvalifikačních prací (VŠKP) obhajovaných na vysokých školách v ČR a pro přenos souborů. Úvodní část zahrnuje cíle, zdůvodnění a východiska vytvoření tohoto národního standardu, dále informace o struktuře popisu zahrnutých prvků, o použitých číselnících, o použitých jmenných prostorech a schématech XML včetně příkladů metadatových záznamů. Hlavní část dokumentu podrobně specifikuje jednotlivé zařazené prvky. V části 2.1 je definováno dvacet čtyři popisných metadatových prvků, v části 2.2 osm prvků technické a administrativní povahy pro potřeby komunikace samotných úplných textů VŠKP.

Status dokumentu

Verze 1.1 z 15. července 2008 popisuje metadatové prvky vysokoškolských kvalifikačních prací. Další verze tohoto standardu mohou tuto verzi nahrazovat. Následující verze bude zveřejněna na stránkách <http://www.evskp.cz/standardy/evskp/>.

Verzi 1.1 z důvodu omezených možností validačního schématu XML ruší atribut `xsi:type` u prvků `dc:identifier` a `dc:rights`. Rozlišení, zda se jedná o URI či volný text, je možno provádět automatizovaně a vypuštěním atributu tak nedojde k omezení funkčnosti ani zásadní změně standardu. Verze 1.1 poprvé obsahuje validační schémata XML a RELAX NG, viz kapitola 1.1 Struktura popisu metadatových prvků, a specifikuje použití prvku `evskp:metadata` nadřazeného prvkům ostatním.

1 Úvod

Tento dokument definuje národní soubor metadatových prvků pro popis vysokoškolských kvalifikačních prací (VŠKP) obhájovaných na vysokých školách v České republice a pro přenos souborů.

Předpokládá se, že tento standard bude výměnným formátem platným pro systémy lokalizované v České republice a také formátem pro výměnu metadat s významnými zahraničními systémy. Standard je doporučeným formátem pro import v Národním registru VŠKP v ČR (NR VŠKP).

Základní skupina uvedených popisných prvků vychází, stejně jako jiné ve světě známé národní metadatové specifikace, především z aktuální verze 1.1 doporučení obecného standardu pro popis informačních zdrojů **Dublin Core** (DCES, 2008-01-14, <http://dublincore.org/documents/dces/>), resp. aktuální verze definic **termínů DCMI** ze 14. ledna 2008 (DCMI-TERMS, 2008-01-14, <http://dublincore.org/documents/dcmi-terms>). Specifikace Dublin Core doznala od doby zveřejnění nulté verze standardu EVSKP-MS 7. 11. 2005 (0.1, <http://www.evskp.cz/standardy/metadaty/navrh-2005-11-07.html>) značných proměn - nejdříve v roce 2006 a následně i v roce 2008. V důsledku toho se první verze standardu EVSKP-MS (1.0) musela v mnoha směrech změnit. Nově byla přijata koncepce tří souběžných standardů: pro popis prací a zvláště pro popis fyzických osob a korporací. Prvky pro popis fyzických osob a korporací mohou být volitelně použity jako prvky vnořené u vybraných prvků popisu VŠKP. Základní popisné prvky a jejich povinné uvádění bylo v zásadě zachováno, přibýly však některé prvky jiné (jako nepovinné). Větší změny uživatelé zaregistrují v oblasti kvalifikujících informací. Standard EVSKP zahrnuje v nové verzi také sadu prvků administrativního a technického charakteru. Jejich přítomnost je nezbytná pro zajištění přenosu samotných souborů s plnými texty mezi registry.

Některé specifické prvky pro popis vysokoškolských kvalifikačních prací byly převzaty z aktuální verze 1.00 (rev. 2) nejvýznamnějšího mezinárodního standardu **ETD-MS** (2008-03-27, <http://www.ndltd.org/standards/metadata/etd-ms-v1.00-rev2.html>), který byl připraven pro potřeby mezinárodního virtuálního systému NDLTD (<http://www.ndltd.org>).

Při tvorbě českého národního standardu bylo dále přihlédnuto z hlediska koncepčního i z hledisek jiných k vybraným národním standardům. Jde zejména o nejnovější verzi 1.3 německého národního metadatového standardu **XMetaDiss** pro detailní popis disertací (2006-08-22, <http://www.d-nb.de/standards/xmetadiss/xmetadiss.htm>). Německý standard je připraven na zcela nové bázi (také v souladu se změnami specifikace Dublin Core), předpokládá komunikaci metadat již jen ve struktuře XML, a u vybraných základních popisných prvků je nabízeno vnořování prvků popisujících fyzické osoby a korporace. Přihlédnuto bylo také ke starší německé specifikaci metadat autoritních entit (fyzických osob) **MetaPers** ve verzi 1.1 (2003-07-07, http://deposit.ddb.de/metadiss.htm#top_II), jejíž nová verze (v XML) doposud není k dispozici. Konzultována byla také aktuální 2. verze návrhu francouzské normy specifikující metadata pro elektronické disertace **TEF 2.0** z března 2006 (2. éd., mars 2006, <http://www.abes.fr/abes/documents/tef/index.html>), zejména pak definice některých prvků. Detailní popisy prvků českého národního standardu také výrazně přihlížejí k praxi popisu VŠKP v ČR.

1.1 Struktura popisu metadatových prvků

Jednotlivé metadatové prvky, prezentované ve druhé kapitole tohoto dokumentu, jsou popsány pomocí tabulkové formy. Prvky jsou v záhlaví uvedeny jménem. V tabulkovém přehledu jsou postupně prezentovány následující údaje:

Označení – pojmenování prvku v českém jazyce

Jméno – pojmenování prvku pro potřeby zápisu v XML

URI – jednoznačný trvalý identifikátor prvku

URI nadřazeného prvku – URI přímo nadřazeného prvku, v rámci kterého je uvedený prvek vnořen. Pokud údaj není uveden, prvek není podřazen žádnému prvu

Povinnost – údaj o povinnosti uvést pokaždé tento prvek v metadatovém záznamu; údaj nabývá hodnot ano (povinný) / ne (volitelný). Tabulka povinného prvku je podbarvena žlutooranžovou barvou. Tabulka volitelného prvku je podbarvena barvou šedou

Opakovatelnost – údaj o možnosti uvést v metadatovém záznamu více výskytů prvku s daným jménem

Atribut povinný – atribut zpřesňující daný prvek, uvedení tohoto atributu v rámci daného prvku je povinné. Hodnoty, jakých může atribut nabývat, jsou uvedeny v hranatých závorkách [], jednotlivé hodnoty jsou odděleny znakem svislá čára |

Atribut volitelný – atribut zpřesňující daný prvek, uvedení tohoto atributu v rámci daného prvku je volitelné. Hodnoty, jakých může atribut nabývat, jsou uvedeny v hranatých závorkách [], jednotlivé hodnoty jsou odděleny znakem svislá čára |. Pokud je v případě neuvedení atributu nějaká hodnota považována za výchozí, je uvedena na prvním místě **tučně**

Definice – textový popis prvku specifikující použití v metadatovém záznamu

Poznámky – doplňující informace určená ke zpřesnění popisu prvku a jeho použití

Příklady – ukázky použití prvku v zápisu XML

Ukázky zápisu prvků v XML jsou zapsány fontem Courier New velikost 9 bodů.

1.2 Jmenné prostory a XML schémata

Jmenný prostor pro Metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR (EVSKP-MS) definovaný v tomto dokumentu je <http://www.evskp.cz/standardy/evskp/>, prefix evskp. Pokud není v tomto dokumentu uveden jiný jmenný prostor prvku, chápe se použití jmenného prostoru evskp.

Nadřazený metadatový prvek evskp:metadata s vnořenými prvky tohoto standardu obsahuje definice jmenných prostorů podle níže uvedeného seznamu.

V tomto standardu jsou použity následující jmenné prostory a XML schémata:

dc

Dublin Core Metadata Element Set, Version 1.1

jmenný prostor: <http://purl.org/dc/elements/1.1/>

XML schéma: <http://dublincore.org/schemas/xmls/qdc/dc.xsd>

dcterms

DCMI Metadata Terms

jmenný prostor: <http://purl.org/dc/terms/>

XML schéma: <http://dublincore.org/schemas/xmls/qdc/dcterms.xsd>

dctype

DCMI Type Vocabulary

jmenný prostor: <http://purl.org/dc/dcmitype/>

XML schéma: <http://dublincore.org/schemas/xmls/qdc/dcmitype.xsd>

ccz

Metadatový soubor pro popis korporací (CorpCZ)

jmenný prostor: <http://www.evskp.cz/standardy/corpcz/>

XML schéma: <http://www.evskp.cz/standardy/corpcz/1.0/ccz.xsd>

RELAX NG schéma (kompaktní syntaxe): <http://www.evskp.cz/standardy/corpcz/1.0/ccz.rnc>

RELAX NG schéma: <http://www.evskp.cz/standardy/corpcz/1.0/ccz.rng>

pcz

Metadatový soubor pro popis fyzických osob (PersCZ)

jmenný prostor: <http://www.evskp.cz/standardy/perscz/>

XML schéma: <http://www.evskp.cz/standardy/perscz/1.0/pcz.xsd>

RELAX NG schéma (kompaktní syntaxe): <http://www.evskp.cz/standardy/perscz/1.0/pcz.rnc>

RELAX NG schéma: <http://www.evskp.cz/standardy/perscz/1.0/pcz.rng>

thesis

ETD-MS: an Interoperability Metadata Standard for Electronic Theses and Dissertations

jmenný prostor: <http://www.ndltd.org/standards/metadata/etdms/1.0/>

evskp

Metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR (tento standard)

jmenný prostor: <http://www.evskp.cz/standardy/evskp/>

XML schéma: <http://www.evskp.cz/standardy/evskp/1.1/evskp.xsd>

RELAX NG schéma (kompaktní syntaxe): <http://www.evskp.cz/standardy/evskp/1.1/evskp.rnc>

RELAX NG schéma: <http://www.evskp.cz/standardy/evskp/1.1/evskp.rng>

1.3 Příklady zápisu

Níže odkazované příklady znázorňují způsob zápisu informací o vysokoškolské kvalifikační práci.

Příklad 1 - minimální

Příklad využívající pouze povinné prvky. Tento způsob zápisu je až na pár výjimek shodný se zápisem podle EVSKP-MS verze 0.1 <http://www.evskp.cz/standardy/metadata/navrh-2005-11-07.html>.

Příklad je na adrese <http://www.evskp.cz/standardy/evskp/1.1/priklad1.htm>

Příklad 2 – popis bez definice přenosu souborů

Příklad využívá většinu prvků popisujících VŠKP. Jsou v něm vynechány specifické prvky pro popis a přenos souborů a přístupových práv, protože podle tohoto příkladu soubory není možné stáhnout.

Příklad je na adrese <http://www.evskp.cz/standardy/evskp/1.1/priklad2.htm>

Příklad 3 – přenos plných textů

Příklad se zaměřuje na využití administrativních a technických prvků, specifických možnosti přenosu souborů a jejich zpřístupnění.

Příklad je na adrese <http://www.evskp.cz/standardy/evskp/1.1/priklad3.htm>

2 Definice prvků

Prvky podle tohoto metadatového standardu jsou uváděny se jmenným prostorem evskp, pokud není uvedeno jinak.

Abecední seznam použitých prvků dc:

dc:creator, dc:contributor, dc:identifier, dc:language, dc:publisher, dc:rights, dc:subject, dc:title, dc:type

Abecední seznam použitých prvků dcterms:

dcterms:abstract, dcterms:available, dcterms:alternative, dcterms:bibliographicCitation, dcterms:created, dcterms:dateAccepted, dcterms:dateSubmitted, dcterms:extent, dcterms:medium, dcterms:modified, dcterms:tableOfContents

Abecední seznam použitých prvků thesis:

thesis:degree, thesis:discipline, thesis:level, thesis:name, thesis:grantor

Abecední seznam definovaných prvků evskp:

evskp:contact, evskp:dateDelivered, evskp:metadata, evskp:modified, evskp:fileNumber, evskp:fileProperties, evskp:server, evskp:transfer

2.1 Popisné metadatové prvky

evskp:metadata

Označení	Popis vysokoškolské kvalifikační práce
Jméno	evskp:metadata
URI	http://www.evskp.cz/terms/evskp/metadata
Povinnost	Ano
Opakovatelnost	Ne
Atribut povinný	version="1.1"
Atribut volitelný	
Definice	Kořenový prvek metadatového záznamu VŠKP, ostatní prvky níže uvedené jsou prvky vnořené.
Poznámky	<p>V rámci kořenového prvku evskp:metadata jsou uváděny jmenné prostory použité ve vnořených prvcích metadatovém záznamu, viz příklad.</p> <p>Atribut version označuje číslo verze standardu EVSKP-MS použitého pro popis vysokoškolské kvalifikační práce. Pokud prvek evskp:metadata neobsahuje atribut version, metadatový záznam je napsán v předchozí verzi standardu.</p>
Příklady	<pre><evskp:metadata version="1.1" xmlns:evskp="http://www.evskp.cz/standards/evskp/"</pre>

	<pre> xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:thesis="http://www.ndltd.org/standards/metadata/etdm s/1.0/" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:pcz="http://www.evskp.cz/standardy/perscz/" xmlns:ccz="http://www.evskp.cz/standardy/corpcz/"> ...</evskp:metadata> </pre>
--	--

dc:title

Označení	Název VŠKP
Jméno	dc:title
URI	http://purl.org/dc/terms/title
Povinnost	Ano
Opakovatelnost	Ano
Atribut povinný	xml:lang="[kód jazyka]"
Atribut volitelný	evskp:typeTranslated="translated"
Definice	Hlavní název VŠKP. Uvádí se ve znění z její titulní stránky. Běžně bývá uveden v jazyce, ve kterém je napsán hlavní text celé VŠKP.
Poznámky	<p>Pomocí volitelného atributu „evskp:typeTranslated“ se uvádí varianta hlavního názvu. Hodnota „translated“ označuje překlad hlavního názvu VŠKP do cizího jazyka, především do angličtiny. U disertačních prací z ČR (prvek dc:type="Disertační práce") je překlad hlavního názvu do angličtiny povinný, u ostatních typů VŠKP je volitelný.</p> <p>Atribut evskp:typeTranslated="translated" se přiděluje také hlavnímu názvu v cizím jazyce, který uvedl autor v prelimináriích česky psaného dokumentu (souběžný název).</p> <p>Pokud je celý text VŠKP napsán cizím jazykem, překlad hlavního názvu do češtiny se zapisuje volitelně s atributem evskp:typeTranslated="translated".</p>
Příklady	<p>1. Na titulní stránce název pouze v češtině</p> <pre> <dc:title xml:lang="cs">Vliv lidské motivace v kolektivních smlouvách na podnikovou sociální politiku</dc:title> <dc:title xml:lang="en" evskp:typeTranslated="translated">Influence of human motivation in collective contracts at the enterprise social policy</dc:title> </pre> <p>2. Na titulní stránce název v češtině a souběžně ve francouzštině</p> <pre> <dc:title xml:lang="cs">Použití souřadnice "hydrostatický tlak" pro integraci elastického modelu dynamiky atmosféry v numerickém předpovědním systému ARPEGE/ALADIN</dc:title> <dc:title xml:lang="fr" evskp:typeTranslated="translated">Utilisation de la </pre>

	<p>coordonnée "pression hydrostatique" pour l'integration des équations élastiques de la dynamique atmosphérique dans le système de prévision numérique ARPEGE/ALADIN</dc:title></p> <p><dc:title xml:lang="en" evskp:typeTranslated="translated">Use of "hydrostatic pressure" coordinate for integration of dynamic atmosphere elastic model in the numeral prevision system ARPEGE/ALADIN</dc:title></p> <p>3. Na titulní stránce anglicky psané práce název pouze v angličtině</p> <p><dc:title xml:lang="en">Dynamics of phase transitions and chemical reactions in semiconductors and semiconductors thin films induced by laser pulse irradiation</dc:title></p>
--	---

dcterms:alternative

Označení	Podnázev VŠKP
Jméno	dcterms:alternative
URI	http://purl.org/dc/terms/alternative
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	xml:lang="[kód jazyka]"
Atribut volitelný	evskp:typeTranslated="translated"
Definice	Podnázev VŠKP nebo další informace k hlavnímu názvu. Uvádějí se ve znění z titulní stránky. Podnázvů nebo dalších informací k hlavnímu názvu může být více. K podnázvům nepatří formální označení typu VŠKP - ten se uvádí v prvku dc:type.
Poznámky	<p>Pomocí volitelného atributu „evskp:typeTranslated“ se uvádí varianta podnázvu. Hodnota "translated" označuje překlad podnázvu VŠKP do cizího jazyka, především do angličtiny.</p> <p>U disertačních prací (prvek dc:type="Disertační práce") obhajovaných v ČR je překlad podnázvu do angličtiny povinný, u ostatních typů VŠKP je volitelný.</p> <p>Atribut evskp:typeTranslated="translated" se přiděluje také podnázvu v cizím jazyce, který uvedl autor v prelimináriích česky psaného dokumentu (souběžný podnázev).</p> <p>Pokud je celý text VŠKP napsán cizím jazykem, překlad podnázvu práce do češtiny se zapisuje volitelně, s atributem evskp:typeTranslated="translated".</p>
Příklady	<p><dc:title xml:lang="cs">Thalamus</dc:title> <dcterms:alternative xml:lang="cs">modulátor normální a patologické aktivity mozku</dcterms:alternative></p> <p><dc:title xml:lang="en" evskp:typeTranslated="translated">Thalamus</dc:title> <dcterms:alternative xml:lang="en"></p>

	evskp:typeTranslated="translated">the modular of normal and pathological activity of brain</dcterms:alternative>
--	--

dc:creator

Označení	Autor VŠKP
Jméno	dc:creator
URI	http://purl.org/dc/terms/creator
Povinnost	Ano
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Fyzická osoba, která je primárně odpovědná za vytvoření obsahu VŠKP, tj. její autor.
Poznámky	<p>Jméno autora se zapisuje dvěma možnými způsoby:</p> <ol style="list-style-type: none"> 1) ve formě: <i>Příjmení, Křestní jméno</i>. Příjmení je od křestního jména odděleno povinně čárkou a mezerou. Příjmení může zahrnovat více částí, oddělených mezerou, pomlčkou apod. Křestní jméno může být po další mezeře následováno dalšími jmény (například <i>Kolínský, Jiří František</i>). Pokud nelze u určitých typů jmen rozlišit příjmení a křestní jméno, zapisuje se v nestrukturované formě. Doplňující informaci pro jednoznačnější identifikaci autora lze uvést po středníku a mezeře (například rok narození, tituly aj.) 2) ve formě prvku pcz:person vnořeného v prvku dc:creator (specifikace prvků pcz:person je uvedena v dokumentu Metadatový soubor pro popis fyzických osob (PersCZ))
Příklady	<p>1. Zápis autora ve formě Příjmení, Křestní jméno</p> <pre><dc:creator>Šonková, Jitka</dc:creator></pre> <p>2. a) Zápis autora se zdvojeným příjmením a datem narození</p> <pre><dc:creator>Beshajová Pelikánová, Ivana; 1971</dc:creator></pre> <p>2. b) Zápis autora se zdvojeným příjmením a datem narození ve formě prvku pcz:person</p> <pre><dc:creator> <pcz:person> <pcz:academicTitleBefore>Ing.</pcz:academicTitleBefore> <pcz:name> <pcz:foreName>Ivana</pcz:foreName> <pcz:surName>Beshajová Pelikánová</pcz:surName> </pcz:name> <pcz:dateOfBirth>1971</pcz:dateOfBirth> </pcz:person> </dc:creator></pre>

	<p>3. a) Zápis nestrukturovaného jména (vietnamské)</p> <pre><dc:creator>Nguyen Thi Ngoc Chan</dc:creator></pre> <p>3. b) Zápis nestrukturovaného jména (vietnamské) ve formě prvku pcz:person</p> <pre><dc:creator> <pcz:person> <pcz:name> <pcz:personEnteredUnderGivenName>Nguyen Thi Ngoc Chan</pcz:personEnteredUnderGivenName> </pcz:name> </pcz:person> </dc:creator></pre>
--	--

dc:subject

Označení	Věcný popis VŠKP
Jméno	dc:subject
URI	http://purl.org/dc/terms/subject
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	xml:lang="[kód jazyka]"
Atribut volitelný	evskp:typeSubject="[evskp:noScheme dcterms:LCSH dcterms:MESH dcterms:UDC dcterms:DDC dcterms:LCC ...]"
Definice	Téma VŠKP vyjádřené buď ve formě volně tvořených klíčových slov (jednoslovných i víceslovných, evskp:typeSubject="evskp:noScheme"), nebo ve formě termínů z některého z řízených předmětových systémů (hesláře, tezaury), popřípadě ve formě třídniců klasifikačních systémů.
Poznámky	<p>Jednotlivé termíny (klíčová slova, předmětová hesla, deskriptory, třídniců) budou oddělovány od předchozích pomocí středníku a mezery. V opakovatelném výskytu lze v případě předmětových systémů zapisovat ekvivalenty v cizích jazycích, přednostně v angličtině.</p> <p>Kódy vybraných řízených předmětových systémů užívaných v ČR jsou uvedeny v autorizovaném číselníku NK ČR (http://www.nkp.cz/pages/page.php3?page=fond_kodovnik.htm).</p> <p>Kódy vybraných významných světových pořadacích systémů jsou uvedeny v seznamu termínů DCMI (http://dublincore.org/documents/dcmi-terms/#H4).</p>
Příklady	<p>1. (volně tvořená klíčová slova)</p> <pre><dc:subject xml:lang="cs">mluvená čeština; morfologie; frekvenční analýza; počítačové zpracování</dc:subject></pre> <pre><dc:subject xml:lang="en">spoken Czech; morphology; frequency analysis; computer</pre>

	<pre>processing</dc:subject></pre> <p>2. (řízené předmětové heslo NK ČR z databáze AUT)</p> <pre><dc:subject xml:lang="cs" evskp:typeSubject="evskp:czenas">uživatelská rozhraní</dc:subject></pre> <pre><dc:subject xml:lang="en" evskp:typeSubject="evskp:czenas">user interfaces</dc:subject></pre> <p>3. (třídník Deweyho desetinné klasifikace)</p> <pre><dc:subject xml:lang="en" evskp:typeSubject="dcterms:DDC">02</dc:subject></pre>
--	---

dcterms:abstract

Označení	Abstrakt VŠKP
Jméno	dcterms:abstrakt
URI	http://purl.org/dc/terms/abstract
Povinnost	Ano
Opakovatelnost	Ano
Atribut povinný	xml:lang="[kód jazyka]"
Atribut volitelný	
Definice	Textový popis, zahrnující detaily vztahující se k tématu pojednávanému ve VŠKP, zejména cíle práce, hypotézy, metody, závěry apod.
Poznámky	V dalším výskytu se u disertačních prací zapisuje povinně také abstrakt v angličtině.
Příklady	<pre><dcterms:abstract xml:lang="cs">Disertační práce ukazuje trendy a současné možnosti integrace informačních zdrojů knihoven, které se snaží zpřístupňovat své tradiční i elektronické fondy kompaktním a unifikovaným způsobem pro co nejširší uživatelskou základnu. Cílem těchto snah za pomoci moderních technologií je poskytovat online služby s využitím celého informačního portfolia, které má knihovna k dispozici, a to 24 hodin denně. Práce shrnuje zkušenosti a nejosvědčenější postupy budování informačních systémů pro integraci informačních zdrojů.</dcterms:abstract></pre> <pre><dcterms:abstract xml:lang="en">Doctoral thesis discusses future trends as well as current means of libraries' information resources integration providing compact and unified access to both, traditional and electronic collections to broad range of users. These modern- technology integration efforts aim to provide 24 hours a day, 7 days a week, online services using the complete information portfolio available to the library. The paper summarizes first-hand experiences and well-proven</pre>

	practices of the development of information systems for information resources integration.</dcterms:abstract>
--	---

dcterms:tableOfContents

Označení	Obsah VŠKP
Jméno	dcterms:tableOfContents
URI	http://purl.org/dc/terms/tableOfContents
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	xml:lang="[kód jazyka]"
Atribut volitelný	
Definice	Uvádí se obsah VŠKP (zpravidla seznam kapitol v hierarchickém uspořádání na libovolném počtu úrovních).
Poznámky	Zapisovat lze obsah v jiných jazycích, je-li také k dispozici. Obsahem může být i popis jednotlivých částí VŠKP u netextových typů prací.
Příklady	<dcterms:tableOfContents xml:lang="cs">1. Terminologický úvod. 2. Metodologická východiska. 2.1 Stávající stav poznání. 2.2 Metody výzkumu. 3. Analýza vybraných informačních systémů. 3.1 Evropské informační systémy. 3.2 Severoamerické informační systémy. 4. Závěry a doporučení.</dcterms:tableOfContents>

dc:publisher

Označení	Instituce archivující a nebo zpřístupňující VŠKP
Jméno	dc:publisher
URI	http://purl.org/dc/terms/publisher
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	xml:lang="[kód jazyka]"
Definice	Instituce odpovědná za uložení, trvalou archivaci a nebo zpřístupnění VŠKP v elektronické formě. Může jít o instituci jako celek (např. Národní technická knihovna, Ostravská univerzita) nebo o její podřízenou jednotku (Univerzitní knihovna Ostravské univerzity).
Poznámky	Zpřístupňující institucí nemusí být instituce (škola), která přidělila akademický titul. Jméno instituce se zapisuje dvěma možnými způsoby:

	<p>1) celé jméno se uvádí v prvku "dc:publisher" jako text; jde-li o podřízenou jednotku, zapisuje se pomocí formalizovaného zápisu: Instituce. Podřízená jednotka</p> <p>2) v prvku ccz:universityOrInstitution vnořeném v prvku dc:publisher (specifikace prvku ccz:universityOrInstitution je uvedena v dokumentu Metadatový soubor pro popis korporací (CorpCZ))</p> <p>V opakovatelných výskytech prvku dc:publisher lze uvést i více institucí. V případě potřeby lze popis instituce uvádět v opakovatelných výskytech v cizích jazycích.</p>
Příklady	<p>1. Zápis jména instituce textem</p> <pre><dc:publisher>Národní technická knihovna</dc:publisher></pre> <p>2. Zápis jména instituce a podřízené jednotky textem</p> <pre><dc:publisher>Vysoká škola ekonomická v Praze. Centrum informačních a knihovnických služeb</dc:publisher></pre> <p>2. Zápis jména instituce a podřízené jednotky ve formě prvku ccz:universityOrInstitution</p> <pre><dc:publisher> <ccz:universityOrInstitution> <ccz:name xml:lang="cs">Akademie múzických umění</ccz:name> <ccz:department> <ccz:name xml:lang="cs">Hudební fakulta</ccz:name> </ccz:department> </ccz:universityOrInstitution> </dc:publisher></pre>

dc:contributor

Označení	Vedoucí nebo oponent VŠKP
Jméno	dc:contributor
URI	http://purl.org/dc/terms/contributor
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	thesis:role="[advisor referee]"
Atribut volitelný	
Definice	Fyzická osoba odpovědná za vedení nebo oponování VŠKP.
Poznámky	<p>Označení role se uvádí jako povinný atribut s hodnotou v anglickém jazyce (vedoucí práce = advisor; oponent = referee).</p> <p>Jméno vedoucího nebo oponenta se zapisuje dvěma možnými způsoby:</p> <p>1) ve formě: <i>Příjmení, Křestní jméno</i>. Příjmení je od křestního jména</p>

	<p>odděleno povinně čárkou a mezerou. Příjmení může zahrnovat více částí, oddělených mezerou, pomlčkou apod. Křestní jméno může být po další mezeře následováno dalšími jmény (například <i>Mareček, Karel Václav</i>). Pokud nelze u určitých typů jmen rozlišit příjmení a křestní jméno, zapisuje se v nestrukturované formě. Doplnující informaci pro jednoznačnější identifikaci osoby lze uvést po středníku a mezeře (například rok narození; identifikátor autoritního záznamu aj.)</p> <p>2) ve formě prvku pcz:person vnořeného v prvku dc:contributor (specifikace prvků pcz:person je uvedena v dokumentu Metadatové schéma pro popis fyzických osob (PersCZ))</p>
Příklady	<p>1. Zápis vedoucího práce ve formě Příjmení, Křestní jméno</p> <pre><dc:contributor thesis:role="advisor">Kotrba, Vladimír; 1945</dc:contributor></pre> <p>2. Zápis oponenta ve formě prvku pcz:person</p> <pre><dc:contributor thesis:role="referee"> <pcz:person> <pcz:academicTitleBefore>prof.</pcz:academicTitleBefore> <pcz:academicTitleBefore>PhDr.</pcz:academicTitleBefore> <pcz:name> <pcz:foreName>Karel</pcz:foreName> <pcz:surName>Bílek</pcz:surName> </pcz:name> <pcz:academicTitleAfter>Ph.D.</pcz:academicTitleAfter> </pcz:person> </dc:contributor></pre>

dcterms:created

Označení	Datum vytvoření VŠKP
Jméno	dcterms:created
URI	http://purl.org/dc/terms/created
Povinnost	Ne
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Datum vytvoření VŠKP. Je zpravidla uvedeno na titulní straně VŠKP.
Poznámky	<p>Údaj o datu vytvoření VŠKP se zapisuje podle standardu W3CDTF (ISO 8601) (http://www.w3.org/TR/NOTE-datetime). Uplatněno může být úplné datum vytvoření VŠKP v modelu RRRR-MM-DD nebo jen rok vytvoření v modelu RRRR.</p> <p>V určitých oborech je datum vytvoření důležité pro potřeby hodnocení oponenta (oponentů) VŠKP.</p>

Příklady	<dcterms:created>2007-11-15</dcterms:created> <dcterms:created>2007</dcterms:created>
----------	--

dcterms:dateSubmitted

Označení	Datum odevzdání či podání VŠKP
Jméno	dcterms:dateSubmitted
URI	http://purl.org/dc/terms/dateSubmitted
Povinnost	Ne
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Datum odevzdání, provedení (u netextových typů prací) či podání (včetně elektronického) VŠKP. Uvádí se, pokud je registrováno počítačovým systémem školy (fakulty, ústavu apod.).
Poznámky	Údaj o datu odevzdání, provedení či podání VŠKP se zapisuje podle standardu W3CDTF (ISO 8601) (http://www.w3.org/TR/NOTE-datetime). Uplatněno může být úplné datum odevzdání či podání VŠKP v modelu RRRR-MM-DD nebo jen rok odevzdání či podání v modelu RRRR.
Příklady	<dcterms:dateSubmitted>2007-12-18</dcterms:dateSubmitted> <dcterms:dateSubmitted>2007</dcterms:dateSubmitted>

dcterms:dateAccepted

Označení	Datum obhajoby VŠKP
Jméno	dcterms:dateAccepted
URI	http://purl.org/dc/terms/dateAccepted
Povinnost	Ano
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Datum obhajoby VŠKP. Zapisuje se přesné datum obhajoby, které je uvedeno také v protokolu obhajoby.
Poznámky	Údaj o datu obhajoby VŠKP se zapisuje podle standardu W3CDTF (ISO 8601) (http://www.w3.org/TR/NOTE-datetime). Uplatněno může být úplné datum obhajoby VŠKP v modelu RRRR-MM-DD nebo jenom

	rok obhajoby v modelu RRRR.
Příklady	<code><dcterms:dateAccepted>2008-01-25</dcterms:dateAccepted></code> <code><dcterms:dateAccepted>2008</dcterms:dateAccepted></code>

dcterms:modified

Označení	Datum modifikace VŠKP
Jméno	dcterms:modified
URI	http://purl.org/dc/terms/modified
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	Datum a popř. čas ukončení procesu modifikace VŠKP (například v rámci migrace souborů za účelem trvalého uložení práce v digitálním úložišti).
Poznámky	Údaj o proběhlé změně souboru VŠKP se zapisuje podle standardu W3CDTF (ISO 8601) (http://www.w3.org/TR/NOTE-datetime). Je preferováno co nepřesnější datum včetně hodin a minut.
Příklady	<code><dcterms:modified></code> <code>2007-12-14T19:20+01:00</code> <code></dcterms:modified></code> <code><dcterms:modified></code> <code>2007-12-14</code> <code></dcterms:modified></code>

dc:type

Označení	Typ VŠKP
Jméno	dc:type
URI	http://purl.org/dc/terms/type
Povinnost	Ano
Opakovatelnost	Ano
Atribut povinný	xml:lang="[kód jazyka]"
Atribut povinný	evskp:typeType="[TypVSKP dcterms:DCMIType]"
Atribut volitelný	
Definice	Označení dílčího typu VŠKP.

Poznámky	<p>Pomocí povinného atributu evskp:typeType se uvádí typ VŠKP. Atribut evskp:typeType="evskp:TypVSKP" označuje hodnoty ze stanoveného číselníku platného v ČR. Povinně se v tomto prvku uvádějí hodnoty v českém jazyce podle národního číselníku evskp:TypVSKP (např. hodnota „Disertační práce“).</p> <p>Atribut evskp:typeType="dcterms:DCMIType" (http://www.dublincore.org/documents/dcmi-terms/#H7) označuje v nepovinných výskytech prvku dc:type hodnoty z obecné typologie informačních zdrojů iniciativy DCMI, tyto hodnoty se uvádějí pouze anglicky. Hodnota "PhysicalObject" označuje metadatový záznam tištěné kvalifikační práce, u které neexistuje elektronická verze (viz příklad 3).</p>
Příklady	<p>1. typologie VŠKP podle národního číselníku evskp:TypVSKP</p> <pre><dc:type xml:lang="cs" evskp:typeType="TypVSKP">Disertační práce</dc:type></pre> <pre><dc:type xml:lang="en" evskp:typeType="TypVSKP">Doctoral thesis</dc:type></pre> <p>2. příklady typologie zdrojů podle DCMI</p> <pre><dc:type xml:lang="en" evskp:typeType="dcterms:DCMIType">Text</dc:type></pre> <pre><dc:type xml:lang="en" evskp:typeType="dcterms:DCMIType">Dataset</dc:type></pre> <pre><dc:type xml:lang="en" evskp:typeType="dcterms:DCMIType">MovingImage</dc:type></pre> <pre><dc:type xml:lang="en" evskp:typeType="dcterms:DCMIType">InteractiveResource</dc:type></pre> <pre><dc:type xml:lang="en" evskp:typeType="dcterms:DCMIType">Sound</dc:type></pre> <pre><dc:type xml:lang="en" evskp:typeType="dcterms:DCMIType">Software</dc:type></pre> <pre><dc:type xml:lang="en" evskp:typeType="dcterms:DCMIType">Service</dc:type></pre> <p>3. příklad záznamu tištěné VŠKP bez elektronické verze</p> <pre><dc:type xml:lang="cs" evskp:typeType="TypVSKP">Disertační práce</dc:type></pre> <pre><dc:type xml:lang="en" evskp:typeType="dcterms:DCMIType">PhysicalObject</dc:type></pre>

dcterms:medium

Označení	Médium (formát souboru) VŠKP
Jméno	dcterms:medium
URI	http://purl.org/dc/terms/medium

Povinnost	Ano
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	Označení datového formátu digitálního souboru hlavní části VŠKP, ve kterém je odevzdána, trvale uložena a/nebo zpřístupňována (například RTF, PDF, LaTeX, HTML aj.).
Poznámky	<p>Hodnoty se zapisují podle mezinárodního autorizovaného číselníku formátů organizace IANA (http://www.iana.org/assignments/media-types/).</p> <p>V případě metadatového záznamu, kdy existuje pouze tištěná VŠKP bez digitálního souboru, uvádí se tento prvek prázdný, bez zadané hodnoty.</p>
Příklady	<pre><dcterms:medium>application/msword</dcterms:medium></pre> <pre><dcterms:medium>application/rtf</dcterms:medium></pre> <pre><dcterms:medium>application/pdf</dcterms:medium></pre> <pre><dcterms:medium>text/html</dcterms:medium></pre> <pre><dcterms:medium>application/zip</dcterms:medium></pre> <pre><dcterms:medium>application/x-compressed</dcterms:medium></pre> <pre><dcterms:medium>image/jpeg</dcterms:medium></pre> <pre><dcterms:medium>audio/mpeg</dcterms:medium></pre> <pre><dcterms:medium>audio/mp4</dcterms:medium></pre>

dcterms:extent

Označení	Rozsah VŠKP
Jméno	dcterms:extent
URI	http://purl.org/dc/terms/extent
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	Rozsah nebo velikost VŠKP.
Poznámky	V případě potřeby se zapisuje přesný rozsah VŠKP (počet stran, příloh

	apod.), délka trvání zvukového záznamu aj.
Příklady	<code><dcterms:extent>320 s., 30 s. příl.</dcterms:extent></code> <code><dcterms:extent>2:50 hod.</dcterms:extent></code> <code><dcterms:extent>15 MB</dcterms:extent></code>

dc:identifier

Označení	Identifikátor VŠKP
Jméno	dc:identifier
URI	http://purl.org/dc/terms/identifier
Povinnost	Ano
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	Jednoznačný identifikátor VŠKP. Může jít o trvalý identifikátor nebo může jít o označení místa uložení VŠKP v elektronické formě, z něhož ji uživatel může získat (například URL).
Poznámky	Hodnota povinného atributu typu identifikátoru zahrnuje libovolné schválené schéma "dcterms:URI" (podle seznamu udržovaného organizací IANA, http://www.iana.org/assignments/uri-schemes.html), například "http:", "urn:nbn", "info:" aj. Při zápisu identifikátorů se postupuje podle doporučení DCMI http://dublincore.org/architecturewiki/ResourceIdentifierGuidelines .
Příklady	<p>1. trvalá URL adresa</p> <pre><dc:identifier> http://www.vse.cz/vskp/eid/5265</dc:identifier></pre> <p>2. identifikátor Handle (hdl) mapovaný do http://</p> <pre><dc:identifier>http://hdl.handle.net/1721.1/14399</dc:identifier></pre> <p>3. a) identifikátor URN:NBN</p> <pre><dc:identifier>urn:nbn:de:bsz:291-scidok-4956</dc:identifier></pre> <p>3. b) identifikátor URN:NBN mapovaný do německé směrovací služby</p> <pre><dc:identifier>http://nbn-resolving.de/urn:nbn:de:bsz:291-scidok-4956</dc:identifier></pre>

dc:language

Označení	Jazyk VŠKP
Jméno	dc:language
URI	http://purl.org/dc/terms/language
Povinnost	Ano
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	Jazyk intelektuálního obsahu VŠKP. Jde o hlavní jazyk, v němž je VŠKP napsána.
Poznámky	<p>Jazyk VŠKP se v povinném atributu zapisuje buď pomocí 2- nebo 3- znakového kódu v souladu s mezinárodní normou ISO 639.</p> <p>Kódy jazyků lze převzít ze seznamu, uveřejněném na webu Registrační autority (Kongresová knihovna) (http://www.loc.gov/standards/iso639-2/).</p> <p>V dalším výskytu lze zaznamenat i jiný jazyk, v němž jsou napsány významné části práce.</p>
Příklady	<dc:language>cs</dc:language>

dcterms:bibliographicCitation

Označení	Bibliografická citace VŠKP
Jméno	dcterms:bibliographicCitation
URI	http://purl.org/dc/terms/extentbibliographicCitation
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	Stručný záznam VŠKP ve formě bibliografické citace.
Poznámky	Zapíše se v souladu s mezinárodním standardem ČSN ISO 690 a ČSN ISO690-2 (s ohledem na české prostředí).
Příklady	<dcterms:bibliographicCitation>KUDRA, František. Integrace finančních trhů. Praha, 2006. 320 s., 30 s. příl. Disertační práce (Ph.D.). Vysoká škola ekonomická. Finanční fakulta</dcterms:bibliographicCitation>

dc:rights

Označení	Práva k využívání VŠKP
Jméno	dc:rights
URI	http://purl.org/dc/terms/rights
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	xml:lang="[kód jazyka]"
Atribut volitelný	
Definice	<p>Textová informace o právech k využívání VŠKP. Specifikují se podmínky, na základě kterých může být práce distribuována, reprodukována atd.; stanovit lze i dobu, po kterou dané podmínky platí a kontakt na osoby, které se o práva starají.</p> <p>Obsahem prvku může volný text nebo URI adresa na WWW stránky s popisem práv.</p>
Poznámky	<p>Obsahem prvku může být přepis licenční smlouvy.</p> <p>Strojově čitelná přístupová práva k jednotlivým souborům jsou řešena v rámci prvků evskp:transfer a dcterms:available.</p>
Příklady	<p>1. Práva k využívání VŠKP jako volný text</p> <pre><dc:rights>Text VŠKP může být volně využíván prostřednictvím Internetu podle licence Creative Commons BY SA.</dc:rights></pre> <p>2. Práva k využívání VŠKP popsána na uvedené URL</p> <pre><dc:rights>http://www.amu.cz/studium/bakalarske-a-magisterske-prace/licencni-smlouva-ve-formatu-pdf-1</dc:rights></pre>

thesis:degree

Označení	Akademický titul nebo vědecko-pedagogická hodnost
Jméno	thesis:degrese
URI	
Povinnost	Ano
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	

Definice	Údaje o akademickém titulu nebo vědecko-pedagogické hodnosti a instituci, která ho přiděluje.
Poznámky	
Příklady	<pre> <thesis:degree> <thesis:name>Ph.D.</thesis:name> <thesis:level>Doktorský</thesis:level> <thesis:discipline>Ekonomie/Ekonomické teorie </thesis:discipline> <thesis:grantor>Vysoká škola ekonomická v Praze. Národohospodářská fakulta</thesis:grantor> </thesis:degree> </pre>

thesis:name (prvek vnořený v prvku thesis:degree)

Označení	Zkratka jména akademického titulu nebo vědecko-pedagogické hodnosti
Jméno	thesis:name
URI	
Povinnost	Ano
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Zkratka jména akademického titulu nebo vědecko-pedagogické hodnosti přidělované vysokou školou (například Ph.D., doc., JUDr. Mgr., Ing. aj.).
Poznámky	Uváděné hodnoty budou přebírány ze stanoveného číselníku a budou v souladu s platným Zákonem o VŠ/1998 (§ 45-47 a 72), popřípadě legislativou platnou v minulosti (pro potřeby retrospektivního zpracování VŠKP).
Příklady	<pre> <thesis:degree> <thesis:name>Ph.D.</thesis:name> ... </thesis:degree> </pre>

thesis:level (prvek vnořený v prvku thesis:degree)

Označení	Typ studijního programu
Jméno	thesis:level
URI	
Povinnost	Ano
Opakovatelnost	Ano

Atribut povinný	xml:lang="[kód jazyka]"
Atribut volitelný	
Definice	Typ studijního programu, v jehož rámci byla připravena VŠKP (doktorský, magisterský, bakalářský aj.).
Poznámky	<p>Uváděné hodnoty budou přebírány ze stanoveného číselníku a budou v souladu s platným Zákonem o VŠ č.111/1998 (§ 45-47 a 72), popřípadě legislativou platnou v minulosti (pro potřeby retrospektivního zpracování VŠKP).</p> <p>Typ studijního programu se uvádí v českém jazyce podle číselníku PASP viz http://founder.uiv.cz/katalog/ciselnik.asp.</p>
Příklady	<pre><thesis:degree> <thesis:level xml:lang="cs">Doktorský</thesis:level> <thesis:level xml:lang="en">Doctoral</thesis:level> ... </thesis/degree></pre>

thesis:discipline (prvek vnořený v prvku thesis:degree)

Označení	Studijní program a studijní obor
Jméno	thesis:discipline
URI	
Povinnost	Ano
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	xml:lang="[kód jazyka]"
Definice	Studijní program a studijní obor akreditovaný příslušnou komisí MŠMT, v jejichž rámci byla připravena a obhájena VŠKP.
Poznámky	<p>Studijní obor se uvádí v českém jazyce podle číselníku AKVO (Studijní obory vysokých škol) na adrese http://founder.uiv.cz/katalog/ciselnik.asp.</p> <p>Seznamy akreditovaných programů a studijních oborů jsou také k dispozici i na WWW serverech jednotlivých škol.</p> <p>Zapíše se dle modelu: Studijní program/Studijní obor</p> <p>Uváděné hodnoty budou přebírány ze stanoveného číselníku.</p>
Příklady	<pre><thesis:degree> <thesis:discipline xml:lang="cs">Ekonomie/Ekonomické teorie </thesis:discipline> ... </thesis/degree></pre>

thesis:grantor (prvek vnořený v prvku thesis:degree)

Označení	Instituce přidávající titul
Jméno	thesis:grantor
URI	
Povinnost	Ano
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	xml:lang="[kód jazyka]"
Definice	Instituce (škola, ústav apod.) garantující přidělení titulu spojené s vytvořením VŠKP.
Poznámky	<p>Jméno instituce se zapisuje dvěma možnými způsoby:</p> <ol style="list-style-type: none"> 1) celé jméno se uvádí v prvku "thesis:grantor"; jde-li o podřízenou jednotku, zapisuje se pomocí formalizovaného zápisu: Instituce. Podřízená jednotka 2) v prvku ccz:universityOrInstitution vnořeném v prvku thesis:grantor (specifikace prvku ccz:universityOrInstitution je uvedena v dokumentu Metadatové schéma pro popis korporací (CorpCZ)) <p>Jména institucí se přebírají ze zákona (veřejné vysoké školy) a statutu škol, příp. z obchodního/živnostenského rejstříku.</p>
Příklady	<ol style="list-style-type: none"> 1. <pre> <thesis:degree> <thesis:grantor xml:lang="cs">Vysoká škola ekonomická v Praze. Národohospodářská fakulta</thesis:grantor> ... </thesis/degree> </pre> 2. <pre> <thesis:degree> <thesis:grantor> <ccz:universityOrInstitution> <ccz:name xml:lang="cs">Vysoká škola ekonomická v Praze</ccz:name> <ccz:department> <ccz:name xml:lang="cs">Národohospodářská fakulta</ccz:name> </ccz:department> </ccz:universityOrInstitution> </thesis:grantor> ... </thesis/degree> </pre>

2.2 Technické a administrativní metadatové prvky

Prvky popisující metadatový záznam, soubory tvořící VŠKP a prvky řešící přenos metadat a těchto souborů do dalších registrů.evskp:contact.

evskp:contact

Označení	Identifikátor poskytovatele metadat
Jméno	evskp:contact
URI	http://www.evskp.cz/terms/evskp/contact
Povinnost	Ano
Opakovatelnost	Ne
Atribut povinný	contactID
Atribut volitelný	
Definice	Identifikátor RID poskytovatele metadat se uvádí v kvalifikátoru contactID ve tvaru předepsaném celostátní matrikou studentů (databáze REGPRO), příp. Ústavem pro informace ve vzdělávání. Identifikuje školu jako celek, fakultu nebo jiné pracoviště školy, na kterém student získal titul. Jméno poskytovatele se uvádí nepovinně jako textový obsah tohoto prvku.
Poznámky	
Příklady	<p>1. Identifikátor poskytovatele metadat – fakulty</p> <pre><evskp:contact contactID="3190"/></pre> <p>2. Identifikátor poskytovatele metadat – fakulty – se slovním popisem</p> <pre><evskp:contact contactID="3190">Vysoká škola ekonomická v Praze. Fakulta informatiky a statistiky</evskp:contact></pre>

evskp:fileNumber

Označení	Počet souborů VŠKP
Jméno	evskp:fileNumber
URI	http://www.evskp.cz/terms/evskp/fileNumber
Povinnost	Ne
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Počet všech dokumentů a příloh tvořících VŠKP, které metadatový záznam popisuje (hlavní práce, přílohy, posudky apod.). Soubory jsou nepovinně popsány v opakováních prvku evskp:fileProperties.
Poznámky	
Příklady	<pre><evskp:fileNumber>2</evskp:fileNumber></pre>

evskp:fileProperties

Označení	Popis konkrétního souboru VŠKP
Jméno	evskp:fileProperties
URI	http://www.evskp.cz/terms/evskp/fileProperties
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	fileID (jednoznačná identifikace prvku evskp:fileProperties, např. přírůstkové číslo)
Atribut volitelný	fileType=[thesis appendix advisorReview refereeReview errata presentation]
Atribut volitelný	fileName (název souboru pro download; povinný při uvedení prvku evskp:transfer)
Atribut volitelný	fileDirectory (adresář se souborem v rámci ZIP archivu; povinný při uvedení prvku evskp:transfer s odkazem na ZIP archiv)
Atribut volitelný	fileSize (velikost souboru v bytech)
Atribut volitelný	format (formát souboru, příp. verze)
Atribut volitelný	characterSet (kódová stránka, ve které je soubor uložen)
Atribut volitelný	creation (název, příp. verze programu, ve kterém byl soubor vytvořen)
Definice	Název souboru je uveden jako obsah prvku, upřesňující specifikace souboru je uvedena v nepovinných kvalifikátorech.
Poznámky	<p>Počet opakování prvku evskp:fileProperties je roven hodnotě prvku evskp:fileNumber, tj. počtu všech souborů VŠKP.</p> <p>V případě, že je použit prvek evskp:transfer, povinně se v prvku evskp:fileProperties uvádí atribut fileName, který odpovídá názvu souboru v prvku evskp:transfer. V případě, že soubory jsou zabaleny pro přenos v archivu ZIP, uvede se povinně atribut directory (directory="/", pokud soubor není v žádném podadresáři archivu).</p>
Příklady	<pre><evskp:fileNumber>4</evskp:fileNumber> <evskp:fileProperties fileName="diplomka.pdf" fileID="soubor001" fileType="thesis" format="PDF1.3" fileSize="4074457" creation="PDFlib.3.03 (WIN32)">Diplomová práce</evskp:fileProperties> <evskp:fileProperties fileName="priloha1.doc" fileID="soubor002" fileType="appendix" format="OpenOffice 1.1" fileSize="1074457" characterSet="UTF-8" fileDirectory="/prilohy/" creation="OpenOffice.org1.1">Příloha 1 Tabulky v OpenOffice</evskp:fileProperties> <evskp:fileProperties fileName="posudek.pdf" fileID="posudek" fileType="advisorReview" format="PDF1.2" fileSize="74457" creation="PDFlib.3.03 (WIN32)" fileDirectory="/posudky/">Hodnocení</pre>

	<pre>vedouciho</evskp:fileProperties> <evskp:fileProperties fileName="oponent.pdf" fileID="oponentural" format="PDF1.2" fileType="refereeReview" fileSize="732457" fileDirectory="/posudky/">Oponentura V. Dvořák</evskp:fileProperties></pre>
--	---

evskp:transfer

Označení	Identifikátor odkazující na soubor tvořící VŠKP nebo archiv ZIP se soubory obsahujícími VŠKP
Jméno	evskp:transfer
URI	http://www.evskp.cz/terms/evskp/transfer
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	accessRights="[domain restrictedAccess public shibboleth eduroam]" (informace o přístupu k souboru: domain=soubor je přístupný pro vyhrazené IP adresy, restrictedAccess=soubor není veřejně přístupný, public=soubor je veřejně přístupný bez omezení, shibboleth=soubor je přístupný pouze po autorizaci protokolem shibboleth, eduroam=soubor je přístupný pouze po autorizaci protokolem eduroam)
Atribut volitelný	fileID (viz prvek evskp:fileProperties; uvádí se, pokud se nejedná o odkaz na archiv ZIP s více soubory)
Definice	Identifikátor URI odkazující na jeden soubor VŠKP s plným textem (prvek evskp:fileNumber=1) nebo archiv souborů komprimovaných formátem ZIP s jedním či více soubory VŠKP (prvek evskp:fileNumber>=1).
Poznámky	<p>Obsahem prvku je URL adresa souboru pro přenos platná v době zpřístupnění metadatového záznamu.</p> <p>Jednotlivé soubory VŠKP jsou popsány v metadatových prvcích evskp:fileProperties.</p> <p>Prvek evskp:transfer se uvádí pouze tehdy, pokud je možné soubor stáhnout do jiného registru mimo primární (lokální) registr, např. za účelem kontroly plagiátů.</p> <p>Soubory na uvedené adrese nemusí být přístupné bez omezení všem registrům. Konkrétní práva pro stažení souboru jsou řešena samostatně (např. povolením IP adresy v lokálním registru, BASIC/DIGEST autentizací, HTTPS přenos ...).</p> <p>Přístup veřejnosti k danému souboru na uvedené adrese je specifikován v povinném atributu accessRights, vzdálený registr musí toto omezení respektovat i tehdy, pokud má přístup k plnému textu (např. pro kontrolu plagiátů). Namísto poskytnutí plného textu např. nabízí URL adresu lokálního registru.</p> <p>V případě použití archivu ZIP s jednotlivými soubory VŠKP se v prvku evskp:transfer neuvádí atribut fileID, v prvcích evskp:fileProperties se</p>

	<p>povinně uvádí atributy fileName a fileDirectory.</p> <p>V případě odkazování na jednotlivé soubory VŠKP se v prvku evskp:transfer povinně uvádí atribut fileID z prvku evskp:fileProperties, v prvcích evskp:fileProperties se povinně uvádí atributy fileName a fileDirectory.</p>
Příklady	<p>1. VŠKP tvořena jedním souborem PDF, lze poskytnout plný text veřejnosti</p> <pre><evskp:transfer evskp:fileID="001" accessRights="public"> http://www.server.cz/files/soubor.pdf </evskp:transfer></pre> <p>2. VŠKP tvořena více soubory popsány v prvcích evskp:fileTransfer, přenos v archivu ZIP, přístup k plnému textu veřejností je omezen v lokálním registru IP adresami</p> <pre><evskp:transfer accessRights="domain">http://www.server.cz/files/archiv.zip</evskp:transfer></pre>

evskp:server

Označení	Informace o serveru zpřístupňujícím VŠKP
Jméno	evskp:server
URI	http://www.evskp.cz/terms/evskp/server
Povinnost	Ne
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Informace o serveru zpřístupňujícím VŠKP, který poskytuje metadata.
Poznámky	Obsahem prvku evskp:server je prvek ccz:universityOrInstitution podle standardu CorpCZ nebo jméno instituce zapsané jako text.
Příklady	<p>1. Jméno instituce zapsané jako text</p> <pre><evskp:server>Vysoká škola ekonomická v Praze. Centrum informačních a knihovnických služeb</evskp:server></pre> <p>2. Jméno instituce zapsané formou prvku ccz:universityOrInstitution</p> <pre><evskp:server> <ccz:universityOrInstitution> <ccz:name xml:lang="cs">Vysoká škola ekonomická v Praze</ccz:name> <ccz:place>Praha</ccz:place> <ccz:department sigla="ABA006"> <ccz:name xml:lang="cs">Centrum informačních a knihovnických služeb</ccz:name> <ccz:email>webmaster@vse.cz</ccz:email> <ccz:homepage>http://ciks.vse.cz/</ccz:homepage> </ccz:department> </ccz:universityOrInstitution> </evskp:server></pre>

evskp:dateDelivered

Označení	Datum doručení metadatového záznamu do repozitáře
Jméno	evskp:dateDelivered
URI	http://www.evskp.cz/terms/evskp/dateDelivered
Povinnost	Ne
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Datum doručení metadatového záznamu do repozitáře z lokálního registru. Vyplňují pouze sekundární repozitáře (např. NR VŠKP), které nejsou primárním zdrojem metadat (tj. nejsou lokálním registrem).
Poznámky	
Příklady	<p>1. Datum doručení s přesností na den</p> <pre><evskp:dateDelivered> 2007-10-15</evskp:dateDelivered></pre> <p>2. Datum doručení s přesností na hodinu a minutu</p> <pre><evskp:dateDelivered> 2004-12-31T19:20+01:00</evskp:dateDelivered></pre>

dcterms:available

Označení	Zpřístupnění souborů VŠKP
Jméno	dcterms:available
URI	http://purl.org/dc/terms/available
Povinnost	Ne
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Datum nebo interval, ve kterém mohou být soubory VŠKP zpřístupněny veřejnosti.
Poznámky	<p>Údaj o zpřístupnění souborů VŠKP se zapisuje podle standardu W3CDTF (ISO 8601) (http://www.w3.org/TR/NOTE-datetime). V případě intervalu se dvě data oddělují znakem lomítko /.</p> <p>Pokud je aktuální datum nižší než datum uvedené v tomto prvku nebo pokud aktuální datum není v intervalu uvedeném v tomto prvku, nesmí</p>

	<p>registr zpřístupnit plné texty ani v případě, že je má k dispozici.</p> <p>Zpřístupnění souborů VŠKP v primárním (lokálním) registru se může řídit zvláštními pravidly, pro tento registr se toto omezení nevztahuje.</p>
Příklady	<p>1. Zpřístupnění povoleno od daného data</p> <pre><dcterms:available>2007-12-18</dcterms:available></pre> <p>2. Zpřístupnění povoleno v daném intervalu</p> <pre><dcterms:available>2007-12-18/2010-31-12</dcterms:available></pre>

evskp:modified

Označení	Datum změny záznamu VŠKP
Jméno	evskp:modified
URI	http://www.evskp.cz/terms/evskp/modified
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	Datum a popř. čas změny metadatového záznamu VŠKP (například modifikace záznamu katalogizátorem knihovny při importu do knihovního katalogu).
Poznámky	Údaj o změně metadatového záznamu se zapisuje podle standardu W3CDTF (ISO 8601) (http://www.w3.org/TR/NOTE-datetime). Je preferováno co nepřesnější datum včetně hodin a minut.
Příklady	<pre><evskp:modified>2007-12-14T19:20+01:00</evskp:modified></pre> <pre><evskp:modified>2007-12-14</evskp:modified></pre>

PersCZ

Metadatový soubor pro popis fyzických osob

Verze 1.0

Tato verze:

<http://www.evskp.cz/standardy/perscz/1.0/>

Aktuální verze:

<http://www.evskp.cz/standardy/perscz/>

Zpracovatelé:

Eva Bratková (Eva.Bratková@ff.cuni.cz)

Jan Mach (machj@vse.cz)

Praha

5. května 2008

Obsah

1	Úvod.....	4
1.1	Struktura popisu metadatových prvků.....	4
1.2	Jmenné prostory	5
1.3	Příklady zápisu	5
2	Definice prvků	7
2.1	Prvek nadřazený	7
	person	7
2.2	Prvky vnořené v prvku person.....	7
	dc:identifier.....	7
	academicTitleBefore	8
	name	9
	foreName (prvek vnořený v prvku name)	9
	surName (prvek vnořený v prvku name).....	10
	personEnteredUnderGivenName (prvek vnořený v prvku name)	10
	academicTitleAfter	11
	dateOfBirth.....	12
	placeOfBirth	12
	note	13
	email	13
	homepage.....	14
	affiliation.....	14

Citace

PersCZ : metadatový soubor pro popis fyzických osob [online]. Zpracovatelé Eva Bratková, Jan Mach. Verze 1.0. Praha : Odborná komise pro otázky elektronického zpřístupňování VŠKP AKVŠ ČR, 2008-05-05 [2008-05-05]. 14 s. Dostupný z WWW: <<http://www.evskp.cz/standardy/perscz/1.0/>>.

Abstrakt

Dokument definuje formální standardní soubor metadatových prvků určených pro popis fyzických osob, které vystupují v roli autorů vysokoškolských kvalifikačních prací (VŠKP) a vedoucích a oponentů těchto prací. Úvodní část zahrnuje cíle, zdůvodnění a východiska vytvoření tohoto standardu, dále informace o struktuře popisu zahrnutých prvků, o použitých jmenných prostorech a schématech XML včetně příkladů metadatových záznamů. Hlavní část dokumentu podrobně specifikuje jednotlivé zařazené prvky. V části 2.1 je definován prvek person, v části 2.2 pak třináct vnořených prvků, určených k popisu vybraných vlastností fyzické osoby (jméno, datum narození, získané akademické tituly aj.).

Status dokumentu

Tato verze z 5. května 2008 je první verzí, která definuje metadatové prvky pro popis fyzických osob. Další verze tohoto standardu mohou tuto verzi nahrazovat. Následující verze bude zveřejněna na stránkách <http://www.evskp.cz/standardy/perscz/>.

1 Úvod

Tento dokument definuje národní soubor metadatových prvků pro popis fyzických osob.

Tvorba tohoto standardu vychází z hlediska koncepčního i z hledisek jiných především z nejnovější verze 1.3 německého národního metadatového standardu **XMetaDiss** pro popis disertací (2006-08-22, <http://www.d-nb.de/standards/xmetadiss/xmetadiss.htm>), který na úrovni vybraných vnořovaných prvků zahrnuje popis fyzických osob. Přihlédnuto bylo také ke starší německé specifikaci metadat pro fyzické osoby **MetaPers** ve verzi 1.1 (2003-07-07, http://deposit.ddb.de/metadiss.htm#top_II), jejíž nová verze (v XML) doposud není k dispozici. Definice prvků se opírají také o vybrané standardy pro popis fyzických osob uplatňovaný v knihovnické katalogizační praxi.

Důvodem pro vytvoření tohoto standardu je potřeba registrace údajů o fyzických osobách ve výměnných metadatových souborech, především tzv. šedé literatury, jako je např. Metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR (EVSKP-MS). V těchto souborech je potřeba počítačově srozumitelným způsobem registrovat údaje o autorech, vedoucích a oponentech prací, konkrétněji jejich jméno, příjmení, tituly, e-mail a další údaje. Formalizace zápisu těchto údajů je uvedena v tomto standardu, ostatní standardy se na tento standard při popisu fyzických osob odvolávají.

Metadatové prvky PersCZ slouží pro popis údajů o fyzické osobě převážně ve výměnných formátech. V lokálních repozitářích může být i nadále udržován aktualizovaný číselník fyzických osob, který lze provázat s popisem v metadatech podle PersCZ pomocí identifikátoru dc:identifier.

1.1 Struktura popisu metadatových prvků

Jednotlivé metadatové prvky, prezentované ve druhé kapitole tohoto dokumentu, jsou popsány pomocí tabulkové formy. Prvky jsou v záhlaví uvedeny jménem. V tabulkovém přehledu jsou postupně prezentovány následující údaje:

Označení – pojmenování prvku v českém jazyce

Jméno – pojmenování prvku pro potřeby zápisu v XML

URI – jednoznačný trvalý identifikátor prvku

URI nadřazeného prvku – URI přímo nadřazeného prvku, v rámci kterého je uvedený prvek vnořen. Pokud údaj není uveden, prvek není podřazen žádnému prvku

Povinnost – údaj o povinnosti uvést pokaždé tento prvek v metadatovém záznamu; údaj nabývá hodnot ano (povinný) / ne (volitelný). Tabulka povinného prvku je podbarvena **žlutooranžovou barvou**. Tabulka volitelného prvku je podbarvena **barvou šedou**

Opakovatelnost – údaj o možnosti uvést v metadatovém záznamu více výskytů prvku s daným jménem

Atribut povinný – atribut zpřesňující daný prvek, uvedení tohoto atributu v rámci daného prvku je povinné. Hodnoty, jakých může atribut nabývat, jsou uvedeny v hranatých závorkách [], jednotlivé hodnoty jsou odděleny znakem svislá čára |

Atribut volitelný – atribut zpřesňující daný prvek, uvedení tohoto atributu v rámci daného prvku je volitelné. Hodnoty, jakých může atribut nabývat, jsou uvedeny v hranatých závorkách [], jednotlivé hodnoty jsou odděleny znakem svislá čára |. Pokud je v případě neuvedení atributu nějaká hodnota považována za výchozí, je uvedena na prvním místě **tučně**

Definice – textový popis prvku specifikující použití v metadatovém záznamu

Poznámky – doplňující informace určená ke zpřesnění popisu prvku a jeho použití

Příklady – ukázky použití prvku v zápisu XML

Ukázky zápisu prvků v XML jsou zapsány fontem Courier New velikost 9 bodů.

1.2 Jmenné prostory a XML schémata

Jmenný prostor pro Metadatové schéma pro popis fyzických osob (PersCZ) definovaný v tomto dokumentu je <http://www.evskp.cz/standardy/perscz/>. Pokud není v tomto dokumentu uveden jiný jmenný prostor prvku, chápe se použití jmenného prostoru PersCZ. V případě použití prvků podle standardu PersCZ v jiném metadatovém standardu se používá výše uvedený jmenný prostor.

V tomto dokumentu jsou použity následující jmenné prostory:

dc

Dublin Core Metadata Element Set, Version 1.1

jmenný prostor: <http://purl.org/dc/elements/1.1/>

XML schéma: <http://dublincore.org/schemas/xmls/qdc/dc.xsd>

ccz

Metadatový soubor pro popis korporací (CorpCZ)

jmenný prostor: <http://www.evskp.cz/standardy/corpcz/>

pcz

Metadatový soubor pro popis fyzických osob (PersCZ)

jmenný prostor: <http://www.evskp.cz/standardy/perscz/>

1.3 Příklady zápisu

Níže uvedený příklad znázorňuje **minimální zápis** údajů o fyzické osobě v rámci jiného metadatového standardu (použit zápis s křestním jménem a příjmením). Pro více informací o dané osobě je možné zápis doplnit o nepovinný prvek dc:identifier odkazující na externí číselník fyzických osob.

```
<pcz:person>
  <pcz:name>
    <pcz:foreName>Martin</pcz:foreName>
    <pcz:surName>Novák</pcz:surName>
  </pcz:name>
</pcz:person>
```

Níže uvedený příklad znázorňuje **rozšířený zápis** informací o fyzické osobě v rámci prvku dc:creator definovaného Metadatovým souborem pro elektronické vysokoškolské kvalifikační práce v ČR (EVSKP-MS). V popisu fyzické osoby je popsána instituce pomocí metadatového souboru CorpCZ, prvku ccz:universityOrInstitution.

```
...<dc:creator>
  <pcz:person>
    <dc:identifier pcz:typeIdentifier="aut">js20061022002</dc:identifier>
    <pcz:academicTitleBefore>doc.</pcz:academicTitleBefore>
    <pcz:academicTitleBefore>Ing.</pcz:academicTitleBefore>
    <pcz:name>
      <pcz:foreName>Martin</pcz:foreName>
      <pcz:surName>Novák</pcz:surName>
    </pcz:name>
    <pcz:academicTitleAfter>CSc.</pcz:academicTitleAfter>,
    <pcz:dateOfBirth>1971-03-23</pcz:dateOf Birth>
    <pcz:placeOfBirth>Brno</pcz:placeOf Birth>
    <pcz:note>Vystudoval Vysokou školu ekonomickou v Praze</pcz:note>
    <pcz:affiliation>
      <ccz:universityOrInstitution>
        <ccz:name xml:lang="cs">Vysoká škola ekonomická v Praze</ccz:name>
        <ccz:place>Praha</ccz:place>
      </ccz:universityOrInstitution>
    </pcz:affiliation>
  </pcz:person>
</dc:creator>
```

```
<ccz:department>
  <ccz:name xml:lang="cs">Fakulta managementu</ccz:name>
  <ccz:place>Jindřichův Hradec</ccz:place>
  <ccz:homepage>http://fm.vse.cz/</ccz:homepage>
</ccz:department>
  <ccz:note>Knihovna FM VŠE JH, veřejně přístupná</ccz:note>
  <ccz:universityOrInstitution>
    </pcz:affiliation>
  </pcz:person>
</dc:creator>...
```

2 Definice prvků

Abecední seznam prvků pcz: academicTitleAfter, academicTitleBefore, affiliation, email, dateOfBirth, foreName, homepage, name, note, person, personEnteredUnderGivenName, placeOfBirth, surName

Abecední seznam prvků dc:
dc:identifier

2.1 Prvek nadřazený

Prvek nadřazený se používá v rámci externích metadat s uvedením jmenného prostoru pcz (<http://www.evskp.cz/standardy/perscz/>). Obsahuje prvky vnořené podle kapitoly 2.2.

Níže uvedené prvky jsou v metadatových záznamech odkazujících se na soubor PersCZ zapisovány v rámci prvku person. Prvky podle tohoto metadatového standardu jsou uváděny se jmenným prostorem pcz, pokud není uvedeno jinak (viz podřízené prvky prvku affiliation).

person

Označení	Popis fyzické osoby
Jméno	person
URI	http://www.evskp.cz/terms/perscz/person
URI nadřazeného prvku	
Povinnost	Ano
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	Popis fyzické osoby.
Poznámky	Prvek je využíván pro popis fyzické osoby v dalších metadatových standardech (např. EVSKP-MS), údaje o fyzické osobě jsou uvedeny v prvcích vnořených, viz 2.2 Prvky vnořené v prvku person.
Příklady	<pre><pcz:person> <pcz:name> <pcz:foreName>Jana</pcz:foreName> <pcz:surName>Nováková</pcz:surName> </pcz:name> </pcz:person></pre>

2.2 Prvky vnořené v prvku person

dc:identifier

Označení	Identifikátor fyzické osoby
Jméno	dc:identifier

URI	http://purl.org/dc/terms/identifier
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	pcz:typeIdentifier="[dcterms:URI aut ...]"
Atribut volitelný	
Definice	Jednoznačný identifikátor fyzické osoby.
Poznámky	Hodnota povinného atributu zahrnuje libovolné schválené schéma "dcterms:URI" (ccz:typeIdentifier="dcterms:URI", podle seznamu udržovaného organizací IANA, http://www.iana.org/assignments/uri-schemes.html) nebo hodnotu typu definovanou specifickým číselníkem (pcz:typeIdentifier="aut" - identifikační číslo záznamu v databázi národních autorit Národní knihovny České republiky).
Příklady	<pre><dc:identifier pcz:typeIdentifier="dcterms:URI"> http://www.vse.cz/persons/novak/</dc:identifier></pre> <pre><dc:identifier pcz:typeIdentifier="aut"> js20061022002</dc:identifier></pre>

academicTitleBefore

Označení	Akademický titul před jménem
Jméno	academicTitleBefore
URI	http://www.evskp.cz/terms/perscz/academicTitleBefore
URI nadřazeného prvku	http://www.evskp.cz/terms/perscz/person
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	Akademické tituly nebo vědecko-pedagogické hodnosti fyzické osoby umístované před jménem.
Poznámky	Tituly se zapisují podle stanoveného číselníku. V případě autora vysokoškolské kvalifikační práce jde o tituly, které autor získal před obhajobou.
Příklady	<pre><academicTitleBefore>doc.</academicTitleBefore> <academicTitleBefore>Ing.</academicTitleBefore> <name> <foreName>Jana</foreName> <surName>Nováková</surName> </name></pre>

name

Označení	Jméno fyzické osoby
Jméno	name
URI	http://www.evskp.cz/terms/perscz/name
URI nadřazeného prvku	http://www.evskp.cz/terms/perscz/person
Povinnost	Ano
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Detailní údaje o jménu fyzické osoby.
Poznámky	<p>Preferovaný zápis je podle občanského průkazu nebo cestovního dokladu.</p> <p>Preferované pořadí zápisu prvků je podle skutečného výskytu při psaní jména, tj. tituly před jménem, křestní jméno, příjmení, tituly za jménem.</p> <p>Povinně se v rámci prvku name uvádějí buď prvky foreName a surname (křestní jméno a příjmení) nebo prvek personEnteredUnderGivenName (pokud se ve jménu neodlišuje křestní jméno a příjmení).</p>
Příklady	<pre><name> <foreName>Martin</foreName> <surName>Novák</surName> </name></pre>

foreName (prvek vnořený v prvku name)

Označení	Křestní jméno fyzické osoby
Jméno	foreName
URI	http://www.evskp.cz/terms/perscz/foreName
URI nadřazeného prvku	http://www.evskp.cz/terms/perscz/name
Povinnost	Ne (Ano, pokud není užít prvek personEnteredUnderGivenName)
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Křestní jméno (jména) fyzické osoby.
Poznámky	Pokud má fyzická osoba kromě křestního jména i další jména (rodové

	jméno aj.), oddělují se od předchozího jména jednou mezerou. Povinně se v rámci prvku name uvádějí buď prvky foreName a surname (křestní jméno a příjmení) nebo prvek personEnteredUnderGivenName (pokud se neodlišuje křestní jméno a příjmení).
Příklady	<pre><name> <foreName>Jitka</foreName> <surName>Šonková</surName> </name></pre>

surName (prvek vnořený v prvku name)

Označení	Příjmení fyzické osoby
Jméno	surName
URI	http://www.evskp.cz/terms/perscz/surName
URI nadřazeného prvku	http://www.evskp.cz/terms/perscz/name
Povinnost	Ne (Ano, pokud není užít prvek personEnteredUnderGivenName)
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Příjmení fyzické osoby.
Poznámky	<p>Příjmení může zahrnovat více částí, které mohou být odděleny mezerami, pomlčkami apod.</p> <p>Povinně se v rámci prvku name uvádějí buď prvky foreName a surname (křestní jméno a příjmení) nebo prvek personEnteredUnderGivenName (pokud se neodlišuje křestní jméno a příjmení).</p>
Příklady	<p>1. Jméno osoby obsahující jedno příjmení</p> <pre><name> <foreName>Jiří Jordan</foreName> <surName>Čermák</surName> </name></pre> <p>2. Jméno osoby obsahující dvě příjmení</p> <pre><name> <foreName>Hana</foreName> <surName>Adámková Heidrová</surName> </name></pre>

personEnteredUnderGivenName (prvek vnořený v prvku name)

Označení	Nestrukturované jméno fyzické osoby
----------	-------------------------------------

Jméno	personEnteredUnderGivenName
URI	http://www.evskp.cz/terms/perscz/personEnteredUnderGivenName
URI nadřazeného prvku	http://www.evskp.cz/terms/perscz/name
Povinnost	Ne (Ano, pokud není užít prvek foreName a zároveň surName)
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Nestrukturovaná forma jména fyzické osoby, kdy nejde rozlišit příjmení a křestní jméno (zejména pro případy vietnamských, korejských aj. fyzických osob)
Poznámky	<p>Jednotlivé části celého nestrukturovaného jména se zpravidla oddělují mezerou. Pokud jde odlišit křestní jméno a příjmení, zapisují se namísto prvku personEnteredUnderGivenName prvky foreName a sureName.</p> <p>Povinně se v rámci prvku name uvádějí buď prvky foreName a surname (křestní jméno a příjmení) nebo prvek personEnteredUnderGivenName (pokud se neodlišuje křestní jméno a příjmení).</p>
Příklady	<pre><name> <personEnteredUnderGivenName>Nguyen Thi Ngoc Chan </personEnteredUnderGivenName> </name></pre>

academicTitleAfter

Označení	Akademický titul za jménem
Jméno	academicTitleAfter
URI	http://www.evskp.cz/terms/perscz/academicTitleAfter
URI nadřazeného prvku	http://www.evskp.cz/terms/perscz/person
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	Akademické tituly fyzické osoby umísťované za jménem.
Poznámky	<p>V případě autora vysokoškolské kvalifikační práce jde o tituly, které autor získal před obhajobou.</p> <p>Příklad uvádí zápis jména „Jana Nováková, MBA, CSc.“.</p>

Příklady	<pre> <name> <foreName>Jana</foreName> <surName>Nováková</surName> </name> <academicTitleAfter>MBA</academicTitleAfter> <academicTitleAfter>CSc.</academicTitleAfter> </pre>
----------	--

dateOfBirth

Označení	Datum narození fyzické osoby
Jméno	dateOfBirth
URI	http://www.evskp.cz/terms/perscz/dateOfBirth
URI nadřazeného prvku	http://www.evskp.cz/terms/perscz/person
Povinnost	Ne
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Datum narození fyzické osoby.
Poznámky	Údaj o datu narození se zapisuje podle standardu W3CDTF (ISO 8601) (http://www.w3.org/TR/NOTE-datetime). Uplatněno může být úplné datum narození v modelu RRRR-MM-DD nebo jenom rok narození v modelu RRRR.
Příklady	<p>1. zápis celého data narození</p> <pre> <dateOfBirth>1971-03-23 </dateOf Birth> </pre> <p>2. zápis roku narození</p> <pre> <dateOfBirth>1971</dateOf Birth> </pre>

placeOfBirth

Označení	Místo narození fyzické osoby
Jméno	placeOfBirth
URI	http://www.evskp.cz/terms/perscz/placeOfBirth
URI nadřazeného prvku	http://www.evskp.cz/terms/perscz/person
Povinnost	Ne
Opakovatelnost	Ne
Atribut povinný	

Atribut volitelný	
Definice	Místo narození fyzické osoby.
Poznámky	Jako místo narození fyzické osoby se zapisuje např. jméno města. Údaj slouží pro bližší identifikaci osoby.
Příklady	<placeOfBirth>Brno</placeOfBirth>

note

Označení	Poznámka o fyzické osobě
Jméno	note
URI	http://www.evskp.cz/terms/perscz/note
URI nadřazeného prvku	http://www.evskp.cz/terms/perscz/person
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	Libovolná poznámka o fyzické osobě.
Poznámky	Poznámka se zapisuje jako volný text.
Příklady	<note>Vystudoval Vysokou školu ekonomickou, obor Informatika.</note>

email

Označení	E-mailová adresa fyzické osoby
Jméno	email
URI	http://www.evskp.cz/terms/perscz/email
URI nadřazeného prvku	http://www.evskp.cz/terms/perscz/person
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	E-mailová adresa fyzické osoby.
Poznámky	

Příklady	<email>martin.dvorak@vse.cz</email>
----------	-------------------------------------

homepage

Označení	WWW adresa domovských stránek fyzické osoby
Jméno	Homepage
URI	http://www.evskp.cz/terms/perscz/homepage
URI nadřazeného prvku	http://www.evskp.cz/terms/perscz/person
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	WWW adresa domovských stránek fyzické osoby.
Poznámky	Pro zápis je použit zápis ve formátu URL.
Příklady	<homepage>http://www.amu.cz/</homepage>

affiliation

Označení	Afiliace
Jméno	affiliation
URI	http://www.evskp.cz/terms/perscz/affiliation
URI nadřazeného prvku	http://www.evskp.cz/terms/perscz/person
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	Instituce, ve kterých či pro které fyzická osoba pracovala či ke kterým náleží, ať již v současnosti nebo v minulosti.
Poznámky	Pro popis instituce se používá prvek ccz:universityOrInstitution podle standardu „Metadatový soubor pro popis korporací (CorpCZ)“ se jmenným prostorem ccz.
Příklady	<pre><affiliation> <ccz:universityOrInstitution> <ccz:name xml:lang="cs">Univerzita Karlova v Praze</ccz:name> <ccz:place>Praha</ccz:place></pre>

	<pre> <ccz:department> <ccz:name xml:lang="cs">Farmaceutická fakulta v Hradci Králové</ccz:name> <ccz:place>Hradec Králové</ccz:place> </ccz:department> <ccz:universityOrInstitution> </affiliation> </pre>
--	--

CorpCZ

Metadatový soubor pro popis korporací

Verze 1.0

Tato verze:

<http://www.evskp.cz/standardy/corpcz/1.0/>

Aktuální verze:

<http://www.evskp.cz/standardy/corpcz/>

Zpracovatelé:

Eva Bratková (Eva.Bratková@ff.cuni.cz)

Jan Mach (machj@vse.cz)

Praha

5. května 2008

Obsah

1	Úvod.....	4
1.1	Struktura popisu metadatových prvků.....	4
1.2	Jmenné prostory	5
1.3	Příklady zápisu	5
2	Definice prvků	6
2.1	Prvek nadřazený	6
	universityOrInstitution	6
2.2	Prvky vnořené v prvku universityOrInstitution	7
	dc:identifier.....	7
	name	7
	place	8
	address	9
	email	9
	homepage.....	10
	note	11
	department.....	11

Citace

CorpCZ : metadatový soubor pro popis korporací [online]. Zpracovatelé Eva Bratková, Jan Mach. Verze 1.0. Praha : Odborná komise pro otázky elektronického zpřístupňování VŠKP AKVŠ ČR, 2008-05-05 [2008-05-05]. 12 s. Dostupný z WWW: <<http://www.evskp.cz/standardy/corpcz/1.0/>>.

Abstrakt

Dokument definuje formální standardní soubor metadatových prvků určených pro popis korporací, které vystupují v roli institucí archivujících a zpřístupňujících vysokoškolské kvalifikační práce (VŠKP) a v roli institucí přidělujících akademické tituly. Úvodní část zahrnuje cíle, zdůvodnění a východiska vytvoření tohoto standardu, dále informace o struktuře popisu zahrnutých prvků, o použitých jmenných prostorech a schématech XML včetně příkladů metadatových záznamů. Hlavní část dokumentu podrobně specifikuje jednotlivé zařazené prvky. V části 2.1 je definován prvek *universityOrInstitution*, v části 2.2 pak osm vnořených prvků, určených k popisu vybraných vlastností korporace (jméno, místo, adresa aj.).

Status dokumentu

Tato verze z 5. května 2008 je první verzí, která popisuje metadatové prvky pro popis korporace. Další verze tohoto standardu mohou tuto verzi nahrazovat. Následující verze bude zveřejněna na stránkách <http://www.evskp.cz/standardy/corpcz/>.

1 Úvod

Tento dokument definuje národní soubor metadatových prvků pro popis korporací, které vystupují v roli institucí archivujících a zpřístupňujících vysokoškolské kvalifikační práce (VŠKP) a v roli institucí přidělujících akademické tituly.

Tvorba tohoto standardu vychází z hlediska koncepčního i z hledisek jiných především z nejnovější verze 1.3 německého národního metadatového standardu **XMetaDiss** pro detailní popis disertací (2006-08-22, <http://www.d-nb.de/standards/xmetadiss/xmetadiss.htm>), který na úrovni vybraných vnořovaných prvků zahrnuje popis korporací. Definice prvků se opírají také o vybrané standardy pro popis korporací uplatňovaný v knihovnické katalogizační praxi.

Předpokládá se, že tento formát bude sloužit pro výměnu dat mezi jednotlivými repositáři v České republice, konkrétně že bude použit v dalších metadatových standardech. Jedním z nich je Metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR (EVSKP-MS).

Metadatové prvky CorpCZ slouží pro popis údajů o korporaci ve výměnných formátech. V lokálních repositářích může být i nadále udržován aktualizovaný číselník korporací, který lze provázat s popisem v metadatach podle CorpCZ pomocí identifikátoru dc:identifier.

Pro potřeby metadatových standardů popisujících šedou literaturu je schéma CorpCZ velmi komplexní. Vzhledem k možnosti opakování jednotlivých prvků a opakovaného vnořování prvku "department" může být reprezentace metadat v informačních systémech komplikovanější. V tomto případě doporučujeme v databázích uchovávat mimo jiné celý záznam korporace ve formátu xml a pro zjednodušení uživateli zobrazovat alespoň:

- první z opakovatelných prvků (např. první uvedené jméno instituce, prvek name)
- obsah nejvíce zanořeného prvku department (pokud je např. v rámci školy uvedena fakulta a v rámci ní katedra, uživateli budou prezentovány minimálně informace o škole a o nejvíce vnořeném prvku - katedře, popis fakulty může být z technických důvodů vynechán)

1.1 Struktura popisu metadatových prvků

Jednotlivé metadatové prvky, prezentované ve druhé kapitole tohoto dokumentu, jsou popsány pomocí tabulkové formy. Prvky jsou v záhlaví uvedeny jménem. V tabulkovém přehledu jsou postupně prezentovány následující údaje:

Označení – pojmenování prvku v českém jazyce

Jméno – pojmenování prvku pro potřeby zápisu v XML

URI – jednoznačný trvalý identifikátor prvku

URI nadřazeného prvku – URI přímo nadřazeného prvku, v rámci kterého je uvedený prvek vnořen. Pokud údaj není uveden, prvek není podřazen žádnému prvku

Povinnost – údaj o povinnosti uvést pokaždé tento prvek v metadatovém záznamu; údaj nabývá hodnot ano (povinný) / ne (volitelný). Tabulka povinného prvku je podbarvena **žlutooranžovou barvou**. Tabulka volitelného prvku je podbarvena barvou šedou

Opakovatelnost – údaj o možnosti uvést v metadatovém záznamu více výskytů prvku s daným jménem

Atribut povinný – atribut zpřesňující daný prvek, uvedení tohoto atributu v rámci daného prvku je povinné. Hodnoty, jakých může atribut nabývat, jsou uvedeny v hranatých závorkách [], jednotlivé hodnoty jsou odděleny znakem svislá čára |

Atribut volitelný – atribut zpřesňující daný prvek, uvedení tohoto atributu v rámci daného prvku je volitelné. Hodnoty, jakých může atribut nabývat, jsou uvedeny v hranatých závorkách [], jednotlivé hodnoty jsou odděleny znakem svislá čára |. Pokud je v případě neuvedení atributu nějaká hodnota považována za výchozí, je uvedena na prvním místě **tučně**

Definice – textový popis prvku specifikující použití v metadatovém záznamu

Poznámky – doplňující informace určená ke zpřesnění popisu prvku a jeho použití

Příklady – ukázky použití prvku v zápisu XML

Ukázky zápisu prvků v XML jsou zapsány fontem Courier New velikost 9 bodů.

1.2 Jmenné prostory a XML schémata

Jmenný prostor pro Metadatové schéma pro popis korporací (CorpCZ) definovaný v tomto dokumentu je <http://www.evskp.cz/standardy/corpcz/>. Pokud není v tomto dokumentu uveden jiný jmenný prostor prvku, chápe se použití jmenného prostoru CorpCZ. V případě použití prvků dle standardu CorpCZ v jiném metadatovém standardu se používá výše uvedený jmenný prostor.

V tomto dokumentu jsou použity následující jmenné prostory:

dc

Dublin Core Metadata Element Set, Version 1.1

jmenný prostor: <http://purl.org/dc/elements/1.1/>

XML schéma: <http://dublincore.org/schemas/xmls/qdc/dc.xsd>

ccz

Metadatový soubor pro popis korporací (CorpCZ)

jmenný prostor: <http://www.evskp.cz/standardy/corpcz/>

1.3 Příklady zápisu

Níže uvedený příklad znázorňuje **minimální zápis** údajů o korporaci v rámci jiného metadatového standardu. Údaje o dané korporaci je možné doplnit o nepovinný prvek dc:identifier odkazující na externí číselník institucí.

```
<ccz:universityOrInstitution>
  <ccz:name>Vysoká škola ekonomická v Praze</ccz:name>
</ccz:universityOrInstitution>
```

Níže uvedený příklad znázorňuje **rozšířený zápis** informací o korporaci – vysoké škole v rámci prvku dc:publisher použitého v Metadatovém souboru pro elektronické vysokoškolské kvalifikační práce v ČR (EVSKP-MS). Fakulta je popsána v prvcích vnořených v prvku ccz:department.

```
...<dc:publisher>
  <ccz:universityOrInstitution>
    <ccz:name>Vysoká škola ekonomická v Praze</ccz:name>
    <dc:identifier ccz:typeIdentifier="IČ">61384399</dc:identifier>
    <ccz:place>Praha</ccz:place>
    <ccz:department>
      <ccz:name>Fakulta managementu</ccz:name>
      <ccz:place>Jindřichův Hradec</ccz:place>
      <ccz:homepage>http://fm.vse.cz/</ccz:homepage>
    </ccz:department>
    <ccz:note>Knihovna FM VŠE</ccz:note>
  </ccz:universityOrInstitution>
</dc:publisher>...
```

2 Definice prvků

Abecední seznam prvků ccz:

address, department, email, homepage, name, note, place, universityOrInstitution

Abecední seznam prvků dc:

dc:identifier

2.1 Prvek nadřazený

Prvek nadřazený se používá v rámci externích metadat s uvedením jmenného prostoru ccz <http://www.evskp.cz/standardy/corpcz/>. Obsahuje prvky vnořené podle kapitoly 2.2.

Níže uvedené prvky jsou v metadatových záznamech odkazujících se na CorpCZ zapisovány v rámci prvku universityOrInstitution. Prvky podle tohoto metadatového standardu jsou uváděny se jmenným prostorem ccz, pokud není uvedeno jinak (viz prvek dc:identifier).

universityOrInstitution

Označení	Popis univerzity nebo instituce
Jméno	universityOrInstitution
URI	http://www.evskp.cz/terms/corpcz/universityOrInstitution
URI nadřazeného prvku	
Povinnost	Ano
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	xml:lang="[cs kód jazyka]"
Atribut volitelný	
Definice	Informace o korporaci – univerzitě nebo jiné instituci a případně jí podřízené jednotce
Poznámky	Prvek je využíván pro popis korporace v externích metadatových standardech (např. EVSKP-MS), údaje o korporaci jsou uvedeny v prvcích vnořených, viz 2.2 Prvky vnořené v prvku universityOrInstitution.
Příklady	<pre><ccz:universityOrInstitution> <ccz:name>Vysoká škola ekonomická v Praze</ccz:name> <dc:identifier ccz:typeIdentifier="IČ">61384399</dc:identifier> <ccz:place>Praha</ccz:place> <ccz:email>webmaster@vse.cz</ccz:email> <ccz:homepage>http://www.vse.cz/</ccz:homepage> <ccz:department> <ccz:name>Národohospodářská fakulta</ccz:name> </ccz:department> </ccz:universityOrInstitution></pre>

2.2 Prvky vnořené v prvku universityOrInstitution

Níže specifikované prvky dc:identifier, name, place, homepage, email a contact mohou být vnořeny v prvku universityOrInstitution nebo v prvku department (viz URI nadřazeného prvku).

dc:identifier

Označení	Identifikátor korporace
Jméno	dc:identifier
URI	http://purl.org/dc/terms/identifier
URI nadřazeného prvku	http://www.evskp.cz/terms/corpcz/universityOrInstitution http://www.evskp.cz/terms/corpcz/department
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	ccz:typeIdentifier="[sigla IČ RID aut dcterms:URI ...]"
Atribut volitelný	
Definice	Jednoznačný identifikátor korporace.
Poznámky	Hodnota povinného atributu zahrnuje libovolné schválené schéma "dcterms:URI" (ccz:typeIdentifier="dcterms:URI", podle seznamu udržovaného organizací IANA, http://www.iana.org/assignments/uri-schemes.html) nebo hodnotu typu definovanou specifickým číselníkem (ccz:typeIdentifier="IČ" – identifikační číslo podnikatele, zápis bez mezer; ccz:typeIdentifier="RID" – ID university nebo fakulty v databázi REGPRO matricy SIMS, příp. rozšířené o identifikaci katedry; ccz:typeIdentifier="aut" - identifikační číslo záznamu v databázi národních autorit Národní knihovny České republiky; ccz:typeIdentifier="sigla" – jednoznačné označení instituce podle systému sigel).
Příklady	<pre><dc:identifier ccz:typeIdentifier="RID">3100</dc:identifier> <dc:identifier ccz:typeIdentifier="IČ">61384399</dc:identifier> <dc:identifier ccz:typeIdentifier="sigla">ABA006</dc:identifier></pre>

name

Označení	Jméno instituce nebo podřízené jednotky
Jméno	name
URI	http://www.evskp.cz/terms/corpcz/name
URI nadřazeného prvku	http://www.evskp.cz/terms/corpcz/universityOrInstitution http://www.evskp.cz/terms/corpcz/department
Povinnost	Ano

Opakovatelnost	Ano
Atribut povinný	xml:lang="[kód jazyka]"
Atribut volitelný	
Atribut volitelný	
Definice	Jméno instituce nebo podřízené jednotky.
Poznámky	V případě, kdy chceme odkázat na konkrétní podřízenou jednotku instituce (fakultu, ústav, katedru, knihovnu apod.), hlavní prvek name obsahuje jméno instituce jako celku a v rámci prvku department je uveden prvek name obsahující jméno podřízené jednotky.
Příklady	<pre> <name xml:lang="cs">Vysoká škola báňská – Technická univerzita Ostrava</name> <name xml:lang="en">VŠB – Technical University of Ostrava</name> <department> <name xml:lang="cs">Hornicko-geologická fakulta</name> </department> </pre>

place

Označení	Místo sídla instituce nebo podřízené jednotky
Jméno	place
URI	http://www.evskp.cz/terms/corpcz/place
URI nadřazeného prvku	http://www.evskp.cz/terms/corpcz/universityOrInstitution http://www.evskp.cz/terms/corpcz/department
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	xml:lang="[cs kód jazyka]"
Definice	Místo sídla instituce nebo podřízené jednotky (město aj.)
Poznámky	<p>Pokud existuje více míst, uvádí se v opakováních prvku. Místo sídla se v cizím jazyce uvádí v opakování prvku s atributem xml:lang.</p> <p>V případě, kdy je nutné odkázat na konkrétní podřízenou jednotku instituce (fakultu, ústav, katedru, knihovnu apod.) a její místo se liší od místa mateřské instituce, hlavní prvek place může obsahovat místo instituce jako celku (např. hlavní, kontaktní adresu) a v rámci prvku department je uveden vnořený prvek place obsahující umístění podřízené jednotky.</p> <p>Konkrétní adresa korporace je uváděna v prvku address.</p>

Příklady	<pre> <place>Praha</place> <place xml:lang="en">Prague</place> <department> <name xml:lang="cs">Fakulta managementu</name> <place>Jindřichův Hradec</place> </department> </pre>
----------	---

address

Označení	Adresa sídla instituce nebo podřízené jednotky
Jméno	address
URI	http://www.evskp.cz/terms/corpcz/address
URI nadřazeného prvku	http://www.evskp.cz/terms/corpcz/universityOrInstitution http://www.evskp.cz/terms/corpcz/department
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	xml:lang="[cs kód jazyka]"
Definice	Adresa sídla instituce nebo podřízené jednotky ve formátu: ulice, PSČ Město. Pokud existuje více adres, uvádějí se v opakováních prvku.
Poznámky	V případě, kdy je nutné odkázat na konkrétní podřízenou jednotku instituce (fakultu, ústav, katedru, knihovnu apod.) a její adresa se liší, hlavní prvek address může obsahovat adresu instituce jako celku (např. hlavní, kontaktní adresu) a v rámci prvku department je uveden vnořený prvek address obsahující adresu podřízené jednotky.
Příklady	<pre> <address>nám. W. Churchilla 4, 130 67 Praha 3</address> <department> <name xml:lang="cs">Fakulta managementu</name> <address>Jarošovská 1117/II, 377 01 Jindřichův Hradec </address> </department> </pre>

email

Označení	E-mail instituce nebo podřízené jednotky
Jméno	email
URI	http://www.evskp.cz/terms/corpcz/email
URI nadřazeného prvku	http://www.evskp.cz/terms/corpcz/universityOrInstitution http://www.evskp.cz/terms/corpcz/department
Povinnost	Ne

Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	Kontaktní e-mail instituce nebo podřízené jednotky.
Poznámky	V případě, kdy je nutné odkázat na konkrétní podřízenou jednotku instituce (fakultu, ústav, katedru, knihovnu apod.) a její kontaktní e-mail je odlišný, hlavní prvek může obsahovat kontaktní e-mail na instituci jako celek a v rámci prvku department je uveden vnořený prvek email obsahující kontaktní e-mailovou adresu podřízené jednotky.
Příklady	<pre> <email>webmaster@vse.cz</email> <department> <name xml:lang="cs">Fakulta managementu</name> <email>webmaster@fm.vse.cz</email> </department> </pre>

homepage

Označení	Domácí stránka instituce nebo podřízené jednotky
Jméno	homepage
URI	http://www.evskp.cz/terms/corpcz/homepage
URI nadřazeného prvku	http://www.evskp.cz/terms/corpcz/universityOrInstitution http://www.evskp.cz/terms/corpcz/department
Povinnost	Ne
Opakovatelnost	Ano
Atribut povinný	
Atribut volitelný	
Definice	Hlavní, domácí WWW stránka instituce nebo podřízené jednotky zapsaná jako URL.
Poznámky	V případě, kdy je nutné odkázat na konkrétní podřízenou jednotku instituce (fakultu, ústav, katedru, knihovnu apod.) a její domácí WWW stránka je odlišná, hlavní prvek homepage může obsahovat adresu WWW stránky instituce jako celku a v rámci prvku department je uveden vnořený prvek homepage obsahující adresu hlavní WWW stránku podřízené jednotky.
Příklady	<pre> <homepage>http://www.vse.cz/</homepage> <department> <name xml:lang="cs">Fakulta managementu</name> <homepage>http://fm.vse.cz</homepage> </department> </pre>

note

Označení	Poznámka
Jméno	note
URI	http://www.evskp.cz/terms/corpcz/note
URI nadřazeného prvku	http://www.evskp.cz/terms/corpcz/universityOrInstitution http://www.evskp.cz/terms/corpcz/department
Povinnost	Ne
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Doplňující informace k instituci nebo její podřízené jednotce.
Poznámky	
Příklady	<note>Instituce přestěhována z Prahy 1 na Prahu 3 v roce 2008</note>

department

Označení	Podřízená jednotka instituce
Jméno	department
URI	http://www.evskp.cz/terms/corpcz/department
URI nadřazeného prvku	http://www.evskp.cz/terms/corpcz/universityOrInstitution http://www.evskp.cz/terms/corpcz/department
Povinnost	Ne
Opakovatelnost	Ne
Atribut povinný	
Atribut volitelný	
Definice	Administrativně podřízená jednotka instituce (například fakulta, ústav, katedra, knihovna aj.).
Poznámky	<p>Obsahem prvku department jsou povinný vnořený prvek name a nepovinné vnořené prvky dc:identifier, place, homepage, email, contact, note a department odpovídající prvkům korporace definovaným v tomto dokumentu.</p> <p>V případě potřeby popsat jednotku instituce podřízenou jiné podřízené jednotce (např. škola – podřízená fakulta – podřízená katedra), použije se opakovaného vnoření prvku department do nadřazeného prvku</p>

	department.
Příklady	<p>1.</p> <pre><department> <name xml:lang="cs">Fakulta managementu</name> </department></pre> <p>2.</p> <pre><department> <name xml:lang="cs">Fakulta managementu</name> <place>Jindřichův Hradec</place> <email>webmaster@fm.vse.cz</email> <homepage>http://fm.vse.cz/</homepage> <department> <name xml:lang="cs">Katedra managementu informací</name> </department> </department></pre>

EVSKP-MS

Metadatový soubor pro elektronické vysokoškolské kvalifikační práce v ČR

Návrh
Verze 0.1
2005-11-07

Tato verze

<http://www.evskp.cz/standardy/metadata/navrh-2005-11-07.html>

Aktuální verze

<http://www.evskp.cz/standardy/metadata/>

Zpracovatelé

Eva Bratková (Eva.Bratkova@ff.cuni.cz)

Eva Bulínová (bulinova@ruk.cuni.cz)

Jan Mach (machj@vse.cz)

Obsah

[1 Úvod](#)

[2 Metadatové prvky a jejich reprezentace](#)

[2.1 Metadatové prvky](#)

[2.1.1 Název VŠKP](#)

[2.1.2 Podnázev VŠKP](#)

[2.1.3 Překlad názvu VŠKP](#)

[2.1.4 Překlad podnázvu VŠKP](#)

[2.1.5 Autor VŠKP](#)

[2.1.6 Datum narození autora VŠKP](#)

[2.1.7 Předmětový popis VŠKP](#)

[2.1.8 Abstrakt VŠKP](#)

[2.1.9 Instituce archivující a zpřístupňující VŠKP](#)

[2.1.10 Osoby podílející se na vedení VŠKP](#)

[2.1.11 Osoby podílející se na oponování VŠKP](#)

[2.1.12 Datum vytvoření / odevzdání či podání VŠKP](#)

[2.1.13 Datum obhájení VŠKP](#)

[2.1.14 Typ VŠKP](#)

[2.1.15 Formát VŠKP](#)

[2.1.16 Identifikátor VŠKP](#)

[2.1.17 Jazyk VŠKP](#)

[2.1.18 Práva k využívání VŠKP](#)

[2.1.19 Přidělovaný akademický titul](#)

[2.1.20 Typ studijního programu](#)

[2.1.21 Studijní program a Studijní obor](#)

[2.1.22 Instituce přidělující titul](#)

[2.2 Reprezentace metadatových prvků](#)

[2.2.1 RDF/XML](#)

[2.2.2 XML](#)

[2.2.3 HTML/XHTML](#)

[2.3 Schémata kódování dat](#)

[2.3.1 Datum](#)

[2.3.2 Jazyk](#)

[2.3.3 Formát zdroje](#)

[2.3.4 Jména fyzických osob](#)

[2.4 Seznamy a číselníky](#)

[2.5 Příklad metadatového záznamu](#)

[3 Příspěvatelé](#)

1 Úvod

Tento dokument definuje standardní národní metadatový soubor pro popis elektronických vysokoškolských kvalifikačních prací (e-VŠKP) obhájovaných na vysokých školách v České republice a metody reprezentace zařazených metadatových prvků pro potřeby jejich komunikace v prostředí počítačových sítí.

Některé vysoké školy v ČR, které již v současnosti zpracovávají ve svých lokálních systémech vysokoškolské kvalifikační práce v elektronické formě, připravily buď vlastní lokálně platné specifikace metadat, nebo adaptovaly specifikace z jiných systémů (domácích nebo zahraničních). Lokální specifikace jsou zpravidla definovány tak, aby splňovaly požadavky kladené pedagogickými nebo vědeckovýzkumnými pracovníky dané školy a také jejich vysokoškolskými knihovníky.

Tento dokument není určen k tomu, aby nahradil jednotlivé stávající lokální standardy metadat. Měl by plnit roli závazného národního standardu v jednotném systému národního registru pro sdílení metadat o e-VŠKP.

Předkládaná verze dokumentu (0.1) je prvním návrhem ze dne 7. listopadu 2005. Odbornou komisí pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací AKVŠ je předkládána širší odborné veřejnosti k posouzení a připomínkování. Veškeré připomínky, náměty nebo dotazy adresujte pomocí elektronické pošty zpracovatelům návrhu (jejich email adresy viz titulní stránka). Zpracované připomínky budou promítnuty do dalších verzí dokumentu.

2 Metadatové prvky a jejich reprezentace

2.1 Metadatové prvky

V níže uvedeném přehledu je definováno celkem 22 metadatových prvků pro popis VŠKP v elektronické formě. Základní skupina uvedených prvků vychází z aktuální verze standardu Dublin Core (<http://dublincore.org/documents/dces/>), některé byly převzaty z aktuální verze standardu ETD-MS (<http://www.ndltd.org/standards/metadata/current.html>), který byl připraven mezinárodním systémem NDLTD (<http://www.ndltd.org/index.en.html>). Zařazení některých kvalifikátorů definovaných metadatových prvků byl inspirován německým národním standardem METADISS (<http://deposit.ddb.de/metadiss.htm>). Detailní popisy prvků českého národního standardu však zcela přihlížejí k praxi popisu VŠKP v ČR.

Jednotlivé metadatové prvky jsou uvedeny českým jménem v titulku, v tabulkovém přehledu je postupně prezentováno označení prvku a definované kvalifikátory, následuje informace o možnosti opakovaného výskytu prvku v jednom záznamu a informace o jeho závaznosti (povinný-nepovinný). Popis zahrnuje detailní informaci o prvku včetně jeho zápisu. Připojené příklady syntaxe v kódu XHTML jsou prozatím doplňkem specifikace – podrobnější informace ke struktuře metadat jsou obsaženy v částech 2.2.-2.3.

2.1.1 Název VŠKP

Označení	DC.title
Kvalifikátor	lang="[kód jazyka]"
Opakovatelnost	Ne
Povinný	Ano
Popis	Hlavní název VŠKP, který uvádí ve znění z její titulní stránky. Je povinný pro všechny typy VŠKP (viz prvek DC.type). Pomocí kvalifikátoru lang se označuje jazyk hlavního názvu.
Syntax XHTML	<meta name="DC.title" xml:lang="[kód jazyka]" content="[Název VŠKP]" />

2.1.2 Podnázev VŠKP

Označení	DC.title.alternative
Kvalifikátor	lang="[kód jazyka]"
Opakovatelnost	Ano
Povinný	Ne
Popis	Podnázev VŠKP nebo další informace k hlavnímu názvu. Uvádějí se ve znění z titulní stránky. Podnázvů nebo dalších informací k hlavnímu názvu může být více. K podnázvům nepatří formální označení typu VŠKP (uvádí se v rámci prvku DC.type). Pomocí kvalifikátoru lang se označuje jazyk podnázvu.
Syntax XHTML	<meta name="DC.title.alternative" xml:lang="[kód jazyka]" content="[Podnázev VŠKP]" />

2.1.3 Překlad názvu VŠKP

Označení	DC.title.translated
Kvalifikátor	lang="[kód jazyka]"
Opakovatelnost	Ano
Povinný	Ne
Popis	Hlavní název VŠKP přeložený do cizího jazyka, v prvním případě do angličtiny. Překlad hlavního názvu do angličtiny je povinný u disertačních prací. V dalších výskytech překladu hlavního názvu lze uplatnit i další jazyky. Překlad hlavního názvu vytváří autor práce a může být uveden již v odevzdané e-VŠKP. Pomocí kvalifikátoru lang se označuje jazyk překladu hlavního názvu.
Syntax XHTML	<meta name="DC.title.translated" xml:lang="[kód jazyka]" content="[Název v cizím jazyce]" />

2.1.4 Překlad podnázvu VŠKP

Označení	DC.title.alternative.translated
Kvalifikátor	lang="[kód jazyka]"
Opakovatelnost	Ano
Povinný	Ne
Popis	Podnázev nebo další informace k názvu přeložené do cizího jazyka, v prvním případě do angličtiny. Překlad podnázvu do angličtiny je povinný u disertačních prací. V dalších výskytech překladu podnázvu lze uplatnit i další jazyky. Překlad podnázvu vytváří autor práce a může být uveden již v odevzdané e-VŠKP. Pomocí kvalifikátoru lang se označuje jazyk přeloženého podnázvu.
Syntax XHTML	<meta name="DC.title.alternative.translated" xml:lang="[kód jazyka]" content="[Podnázev v cizím jazyce]" />

2.1.5 Autor VŠKP

Označení	DC.creator
Kvalifikátor	
Opakovatelnost	Ne
Povinný	Ano
Popis	Entita (osoba) primárně odpovědná za vytvoření obsahu VŠKP, tj. její autor. Jméno se zapisuje ve formě Příjmení, Křestní jméno. Křestní jméno může být po mezeře následováno dalším jménem (například Kolínský, Jiří František).
Syntax XHTML	<meta name="DC.creator" content="[Příjmení, Křestní jméno]" />

2.1.6 Datum narození autora VŠKP

Označení	DC.creator.dateofbirth
Kvalifikátor	scheme="DCTERMS.W3CDTF"
Opakovatelnost	Ne
Povinný	Ne
Popis	Datum narození autora VŠKP. Údaje zapisuje autor práce a mohou být uvedeny také v jejích preliniářích. Údaje o datu narození se zapisují v modelech definovaných standardem konsorcia W3C, který zahrnuje také mezinárodní normu ISO 8601 (http://www.w3.org/TR/NOTE-datetime). Uplatněno může být úplné datum narození v modelu RRRR-MM-DD nebo jenom rok narození v modelu RRRR.
Syntax XHTML	<meta name="DC.creator.dateofbirth" scheme="DCTERMS.W3CDTF" content="RRRR-MM-DD" />
Schéma	<link rel="DCTERMS.W3CDTF" href=" http://www.w3.org/TR/NOTE-datetime " />

2.1.7 Předmětový popis VŠKP

Označení	DC.subject
Kvalifikátor	lang="[kód jazyka]"; scheme="[kód systému]"
Opakovatelnost	Ano
Povinný	Ne
Popis	Téma VŠKP vyjádřené buď ve formě autorských volně tvořených klíčových slov jednoslovných i víceslovných, nebo ve formě termínů z nějakého řízeného předmětového systému (hesláře, tezauru). Kódy vybraných řízených systémů užívaných v ČR jsou uvedeny v autorizovaném číselníku NK ČR (http://www.nkp.cz/pages/page.php3?page=fond_kodovnik.htm). Jednotlivé termíny (klíčová slova, hesla, deskriptory) budou oddělovány od předchozích pomocí středníku. V opakovatelném výskytu lze zapisovat ekvivalenty v cizích jazycích, prioritně v angličtině.
Syntax XHTML	<meta name="DC.subject" xml:lang="[kód jazyka]" content="[klíčová slova]" /> <meta name="DC.subject" scheme="[označení řízeného systému]" xml:lang="[kód jazyka]" content="[termíny z řízeného systému]" />

2.1.8 Abstrakt VŠKP

Označení	DC.description
Kvalifikátor	lang="[kód jazyka]"
Opakovatelnost	Ano
Povinný	Ano
Popis	Textový popis VŠKP, zahrnující detaily vztahující se k tématu, tj. cíle, hypotézy, metody, závěry apod. Text vytváří autor. V dalším výskytu se u disertačních prací zapisuje povinně také abstrakt v angličtině.
Syntax XHTML	<meta name="DC.description" xml:lang="[kód jazyka]" content="[Abstrakt VŠKP]" />

2.1.9 Instituce archivující a zpřístupňující VŠKP

Označení	DC.publisher
Kvalifikátor	
Opakovatelnost	Ano
Povinný	Ne

Popis	Instituce odpovědná za uložení, trvalou archivaci a zejména zpřístupnění VŠKP v elektronické formě. V opakovatelných výskytech lze uvést i více institucí. Zpřístupňující institucí nemusí být instituce (škola), která přidělila akademickou hodnost (titul). Jméno se zapisuje ve formalizované podobě (Univerzita. Fakulta).
Syntax XHTML	<meta name="DC.publisher" content="[Jméno instituce]" />

2.1.10 Osoby podílející se na vedení VŠKP

Označení	DC.contributor.advisor
Kvalifikátor	
Opakovatelnost	Ano
Povinný	Ne
Popis	Entity (osoby) odpovědné za vedení VŠKP. Označení role se v kvalifikující informaci prvku uvádí v anglickém jazyce (advisor). Jména osob se zapisují ve formě: Příjmení, Křestní jméno; doplňující informace. Doplňující informaci lze uvést pro lepší identifikaci entity (rok narození aj.).
Syntax XTML	<meta name="DC.contributor.advisor" content="[Příjmení, Křestní jméno]" />

2.1.11 Osoby podílející se na oponování VŠKP

Označení	DC.contributor.referee
Kvalifikátor	
Opakovatelnost	Ano
Povinný	Ne
Popis	Entity (osoby) odpovědné za oponování VŠKP. Označení rolí se v kvalifikující informaci prvku uvádí v anglickém jazyce (referee). Jména osob se zapisují ve formě Příjmení, Křestní jméno; doplňující informace. Doplňující informaci lze uvést pro lepší identifikaci entity (rok narození aj.).
Syntax XTML	<meta name="DC.contributor.referee" content="[Příjmení, Křestní jméno]" />

2.1.12 Datum vytvoření / odevzdání či podání VŠKP

Označení	DC.date.created
Kvalifikátor	scheme="DCTERMS.W3CDTF"
Opakovatelnost	Ne
Povinný	Ano
Popis	Datum vytvoření VŠKP. Ve většině případů jde zároveň o datum odevzdání práce. Datum se vyskytuje na titulní stránce. Údaje o datu se zapisují v modelech daných standardem konsorcia W3C, který zahrnuje také mezinárodní normu ISO 8601 (http://www.w3.org/TR/NOTE-datetime). Uplatněno může být datum v modelu RRRR-MM-DD, RRRR-MM nebo jen RRRR. Například: 2005-05-12, 2005-05 nebo 2005.
Syntax XHTML	<meta name="DC.date.created" scheme="DCTERMS.W3CDTF" content="[RRRR-MM-DD]" />
Schéma	<link rel="DCTERMS.W3CDTF" href=" http://www.w3.org/TR/NOTE-datetime " />

2.1.13 Datum obhájení VŠKP

Označení	DC.date.accepted
Kvalifikátor	scheme="DCTERMS.W3CDTF"
Opakovatelnost	Ne
Povinný	Ne
Popis	Datum obhajoby VŠKP. Údaj by měl být v podmínkách ČR zapisován až po obhajobě a měl by být autorizován odpovědným pracovníkem komise pro obhajoby. Zapisuje se přesné datum obhajoby, které je uvedeno také v protokolu obhajoby. Údaje o datu se zapisují v modelech definovaných standardem konsorcia W3C, který zahrnuje také mezinárodní normu ISO 8601 (http://www.w3.org/TR/NOTE-datetime). Uplatněno může být datum v modelu RRRR-MM-DD, RRRR-MM nebo jen RRRR. Například: 2005-05-12, 2005-05 nebo 2005.
Syntax XHTML	<meta name="DC.date.accepted" scheme="DCTERMS.W3CDTF" content="[RRRR-MM-DD]" />
Schéma	<link rel="DCTERMS.W3CDTF" href=" http://www.w3.org/TR/NOTE-datetime " />

2.1.14 Typ VŠKP

Označení	DC.type
Kvalifikátor	lang="[kód jazyka]" ; scheme="typ EVSKP"
Opakovatelnost	Ano
Povinný	Ano
Popis	Jde o označení povahy, obsahu či žánru VŠKP. Uvádějí se hodnoty podle schématu EVSKP. V opakovatelném výskytu lze zapisovat hodnoty v jiných jazycích.
Syntax XHTML	<meta name="DC.type" content="disertační práce" /> <meta name="DC.type" content="text.dissertation" /> <meta name="DC.type" content="text.thesis" />

2.1.15 Formát VŠKP

Označení	DC.format
Kvalifikátor	scheme="DCTERMS.IMT"
Opakovatelnost	Ano
Povinný	Ano
Popis	Označení datového formátu digitální kopie VŠKP, ve kterém je odevzdávána, ukládána a/nebo zpřístupňována (například RTF, PDF, LaTeX, HTML aj.). Hodnoty budou předdefinovány podle autorizovaného seznamu formátů organizace IANA (Internet Assigned Numbers Authority) (http://www.iana.org/assignments/media-types/). Seznam formátů je připraven na základě standardů MIME (IMT) - RFC 2045 (http://www.isi.edu/in-notes/rfc2045.txt) a RFC 2046 (http://www.isi.edu/in-notes/rfc2046.txt).
Syntax XHTML	<meta name="DC.format" scheme="DCTERMS.IMT" CONTENT="[Prvek ze seznamu hodnot]" />
Schéma	<link rel="schema.DCTERMS" href=" http://purl.org/dc/terms/ " />

2.1.16 Identifikátor VŠKP

Označení	DC.identifier
Kvalifikátor	scheme="URI"

Opakovatelnost	Ano
Povinný	Ano
Popis	Jednoznačná identifikace e-VŠKP. Může jít o označení místa uložení VŠKP v elektronické formě, z něhož ji uživatel může získat (například URL). URL bude povinně uváděným identifikátorem. Perspektivně je i pro ČR doporučováno zavedení a užívání jedinečných trvalých identifikátorů typu URN, PURL (http://purl.org), Handles (http://handle.net), DOI aj.
Syntax XHTML	<pre><meta name="DC.identifier" scheme="URL" content="[URL e-VŠKP]" /> <meta name="DC.identifier" scheme="URN" content="[URN VŠKP]" /> <meta name="DC.identifier" scheme="URN:NBN:CZ" content="[URN v daném jmenném prostoru NBN v rámci NK ČR webarchivu]" /> <meta name="DC.identifier" scheme="DOI" content="[DOI VŠKP]" /></pre>

2.1.17 Jazyk VŠKP

Označení	DC.language
Kvalifikátor	
Opakovatelnost	Ano
Povinný	Ano
Popis	Jazyk intelektuálního obsahu VŠKP. Jde o hlavní jazyk napsání práce. V dalším výskytu lze zaznamenat i jiný jazyk, v němž jsou napsány významné části práce.
Syntax HTML	<meta name="DC.language" content="[kód jazyka]" />

2.1.18 Práva k využívání VŠKP

Označení	DC.rights
Kvalifikátor	Lang="[kód jazyka]"
Opakovatelnost	Ano
Povinný	Ne
Popis	Textová informace o právech k využívání VŠKP; specifikují se podmínky, na základě kterých může být práce distribuována, reprodukována atd.; stanovit lze i dobu, po kterou dané podmínky platí a kontakt na osoby, které se o práva starají.
Syntax XHTML	<meta name="DC.rights" xml:lang="[kód jazyka]" content="[Úplný text právního významu]" />

2.1.19 Přidělování akademický titul

Označení	thesis.degree.name
Kvalifikátor	
Opakovatelnost	Ne
Povinný	Ano
Popis	Zkratka akademického titulu přidělovaného vysokou školou (například Ph.D., Doc., JUDr. Mgr., Ing. aj.). Údaje budou přebírány z předdefinovaného seznamu a budou v souladu s platným Zákonem o VŠ/1998 (§ 45-47 a 72).
Syntax HTML	<meta name="thesis.degree.name" content="[Zkratka hodnosti]" />

2.1.20 Typ studijního programu

Označení	thesis.degree.level
Kvalifikátor	
Opakovatelnost	Ne
Povinný	Ne
Popis	Typ studijního programu, v jehož rámci byla připravena VŠKP (doktorský, magisterský, bakalářský aj.). Údaje budou uváděny v souladu s platným Zákonem o VŠ/1998 (§ 45-47 a 72)
Syntax XHTML	<meta name="thesis.degree.level" content="[Jméno programu]" />

2.1.21 Studijní program a Studijní obor

Označení	Thesis.degree.discipline
Kvalifikátor	
Opakovatelnost	Ne
Povinný	Ano
Popis	Studijní program a studijní obor akreditovaný příslušnou komisí MŠMT, v jejichž rámci byla připravena a obhájena VŠKP. Seznam studijních programů akreditovaných v ČR je k dispozici na WWW serveru MŠMT: http://www.msmt.cz/files/xls/CelkovyprehledASpr23082005.xls . Seznamy akreditovaných programů a studijních oborů jsou také k dispozici i na WWW seveřech jednotlivých škol. Údaje by se měly uvádět přímo v dokumentu na titulní stránce. Jejich registrace je žádoucí z důvodů efektivního online vyhledávání. Doporučuje se zápis dle modelu: Studijní program/Studijní obor Příklady: Filozofie/Religionistika Ekonomie/Ekonomické teorie
Syntax XHTML	<meta name="thesis.degree.discipline" content="[Jméno Studijní programu/Studijní oboru]" />

2.1.22 Instituce přidělující titul

Označení	thesis.degree.grantor
Kvalifikátor	
Opakovatelnost	Ne
Povinný	Ano
Popis	Instituce (škola) garantující přidělení titulu spojeného s vytvořením VŠKP (například Univerzita Karlova v Praze, Vysoká škola ekonomická aj.). Jméno se zapisuje ve formalizované podobě (Univerzita. Fakulta).
Syntax XHTML	<meta name="thesis.degree.grantor" content="[Jméno instituce]" />

2.2 Reprezentace metadatových prvků

Pro reprezentaci sady metadatových prvků eVŠKP lze použít několik různých syntaxí zápisu, které jsou uvedeny jako doporučení, příp. návrh doporučení Dublin Core Metadata Initiative (DCMI, viz <http://www.dublincore.net/resources/expressions/>). Jedná se o syntaxe:

- RDF/XML
- XML
- HTML/XHTML

Centrální registr odpovědný za sběr metadat musí na straně importu podporovat všechny výše uvedené syntaxe zápisu. Centrální registr bude dále nabízet webové služby, které umožní jednoduchý převod z jednoho podporovaného formátu do druhého.

Lokální systémy, které exportují data do centrálního registru, musí podporovat alespoň jeden z výše jmenovaných způsobů zápisu na straně exportu. Interní reprezentace metadat v lokálních systémech není nijak předepisována a je plně na tvůrcích každého jednotlivého systému.

Je zapotřebí vypracovat systém ověřování metadatových souborů, pro soubory na bázi XML připadá v úvahu starší systém DTD (nedoporučováno používat) a novější XML Scheme, které má rozšířené možnosti validace dat.

2.2.1 RDF/XML

Výchozí standard Resource Description Framework (RDF) je definován konsorciem W3C (viz <http://www.w3.org/RDF/>) a pro potřeby Dublin Core (DC) je specifikován jako jednoduchý (<http://dublincore.org/documents/dcmes-xml/>) nebo kvalifikovaný (<http://dublincore.org/documents/dcq-rdf-xml/>) zápis Dublin Core v RDF/XML. Zápis kvalifikovaného Dublin Core je v přípravě, a proto vycházíme především ze zápisu jednoduchého Dublin Core v RDF/XML.

Dokument musí začínat na prvním řádku deklarací XML

```
<?xml version="1.0"?>
```

Pokud není uvedeno kódování, je použito kódování UTF-8. Mezi doporučená kódování patří UTF-8, ISO-8859-2 a CP-1250.

Po deklaraci následuje reference na XML DTD nebo XML Scheme.

Další část XML dokumentu je tvořena deklarací RDF a jmenných prostorů, např.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/">
```

Uvnitř elementu rdf:RDF je vnořen minimálně jeden element popisující jednu práci, např.

```
<rdf:Description rdf:about="http://example.org/">
...
</rdf:Description>
```

Uvnitř elementu rdf:Description jsou uvedeny jednotlivé prvky popisující e-VŠKP, které mohou obsahovat atributy xml:lang pro popis jazyka nebo rdf:about pro upřesnění metadatového prvku.

Příklad zápisu v RDF/XML

```
<?xml version="1.0" encoding="utf-8" ?> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:thesis="http://eVSKP/scheme/thesis/">
<rdf:Description>
  <dc:type>Diplomová práce</dc:type>
  <dc:title xml:lang="cs" >Úvod do problematiky veřejného výpůjčního práva</dc:title>
  <dc:language>cs</dc:language>
  <dc:creator>Kábrtová, Jana</dc:creator>
  <dc:contributor.advisor>Příbylová, Daniela</dc:contributor.advisor>
  <dc:contributor.referee>Kera, Denisa</dc:contributor.referee>
  <dc:date.created>2001-10-04</dc:date.created>
  <thesis:degree.name>Mgr.</thesis:degree.name>
  <thesis:degree.level>magisterský</thesis:degree.level>
  <thesis:degree.grantor>Univerzita Karlova v Praze. Filozofická fakulta. Ústav informačních studií a knihovnictví</thesis:degree.grantor>
  <dc:publisher>KSV Jinonice</dc:publisher>
  <dc:identifier>
    <dcterms:URI>http://doi.vskp.cz/120.12/31</dcterms:URI>
  </dc:identifier>
```



```
<dc:description>Práce se zabývá problematikou veřejného výpůjčního práva (dále VVP) jak v
historickém kontextu, tak jeho současnou praxí v existujících světových systémech. Práce je rozdělena
do sedmi kapitol. -- První kapitola je věnována světové terminologii užívané v oblasti VVP...
</dc:description>
</rdf:Description>
</rdf:RDF>
```

2.2.2 XML

Standard XML je pro potřeby standardu Dublin Core podrobněji specifikován ve zvláštním doporučení (viz <http://dublincore.org/documents/dc-xml-guidelines/>). Toto doporučení kromě syntaxe pro prvky Dublin Core obsahuje návod pro implementaci prvků nad rámec Dublin Core. Na rozdíl od implementace jednoduchého Dublin Core bude XML obsahovat další prvky definované níže v tomto dokumentu (viz kapitola 5 Kvalifikovaný Dublin Core ve výše zmíněném doporučení).

Základní pravidla kódování metadat:

Pro potřeby eVŠKP bude definováno XML schéma pro validaci dokumentu, použití XML DTD není doporučováno

Pro potřeby eVŠKP budou definovány jmenné prostory pro použití v XML

Jednotlivé kvalifikátory jsou implementovány jako XML elementy a jejich hodnota jako obsah elementů. Použitý jazyk je uveden jako atribut *xml:lang*, použité kódovací schéma jako atribut *xsi:type*

Všechna jména elementů jsou psány malým písmenem

Pro zapsání více metadatových prvků stejného jména je použito opakování XML elementu, příp. s odlišením jazyka a schématu v attributech, např.

```
<dc:title xml:lang="cz">Titulek</dc:title>
<dc:title xml:lang="en">Title</dc:title>
```

Ke zvažení je zápis „rozšířených“ Dublin Core prvků, v úvahu přichází varianty:

```
1. <dc:contributor.advisor>
    Příbylová, Daniela
  </dc:contributor.advisor>
2. <dc:contributor>
    <advisor>Příbylová, Daniela</advisor>
  </dc:contributor>
```

Příklad zápisu v XML:

```
<?xml version="1.0" encoding="utf-8" ?>
<metadata xmlns="http://eVSKP.cz/scheme/" xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:thesis="http://eVSKP.cz/scheme/thesis" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://eVSKP.cz/scheme/ http://eVSKP.cz/scheme/schema.xsd">
  <dc:type>Diplomová práce</dc:type>
  <dc:title lang="cs">Úvod do problematiky veřejného výpůjčního práva</dc:title>
  <dc:language xsi:type="dcterms:RFC3066">cs</dc:language>
  <dc:creator>Kábrtová, Jana</dc:creator>
  <dc:contributor.advisor>Příbylová, Daniela</dc:contributor.advisor>
  <dc:contributor.referee>Kera, Denisa</dc:contributor.referee>
  <dc:date.created xsi:type="dcterms:w3cdtf">2001-10-04</dc:date.created>
  <thesis:degree.name>Mgr.</thesis:degree.name>
  <thesis:degree.level>magisterský</thesis:degree.level>
  <thesis:degree.grantor>Univerzita Karlova v Praze. Filozofická fakulta. Ústav informačních studií a
knihovnictví</thesis:degree.grantor>
  <dc:publisher>KSV Jinonice</dc:publisher>
  <dc:identifier xsi:type="dcterms:URI">http://doi.vskp.cz/120.12/31</dc:identifier>
  <dc:description lang="cs">Práce se zabývá problematikou veřejného výpůjčního práva (dále VVP) jak v
```

```
historickém kontextu, tak jeho současnou praxí v existujících světových systémech. Práce je rozdělena
do sedmi kapitol. -- První kapitola je věnována světové terminologii užívané v oblasti
VVP...</dc:description>
</metadata>
```

2.2.3 HTML/XHTML

Kódování v HTML/XHTML je vhodné pro účely kódování metadat do HTML/XHTML stránky s obsahem e-VŠKP.

Jednotlivé kvalifikátory jsou kódovány jako XHTML značka META,

- atribut *name* obsahuje jméno kvalifikátoru
- atribut *content* obsahuje hodnotu (obsah) kvalifikátoru
- atribut *lang* může obsahovat upřesnění použitého jazyku
- atribut *scheme* může obsahovat upřesnění použitého schématu zápisu

Kvalifikátory DC aj. jsou psány velkými znaky, následuje tečka, pokračování kvalifikátoru může být zapsáno velkými či malými písmeny, první písmeno však musí být vždy malé (jedná se o doporučení, není závazné).
Příklad:

```
<meta name="DC.title" xml:lang="cs" content="Ústavní právo">
<meta name="DC.language" scheme="DCTERMS:RFC3066" content="cs">
```

Příklad zápisu v HTML/XHTML:

```
...
<head profile="http://dublincore.org/documents/dcq-html/">
<title>Ukázka zápisu v HTML/XHTML</title>
<link rel="schema.DC" URL="http://purl.org/dc/elements/1.1/">
<link rel="schema.DCTERMS" URL="http://purl.org/dc/terms/">
<meta name="DC.type" content="Diplomová práce">
<meta name="DC.title" xml:lang="cs" content="Úvod do problematiky veřejného výpůjčního práva">
<meta name="DC.language" scheme="DCTERMS:RFC3066" content="cs">
<meta name="DC.creator" content="Kábrtová, Jana">
<meta name="DC.contributor.advisor" content="Příbylová, Daniela">
<meta name="DC.contributor.referee" content="Kera, Denisa">
<meta name="DC.date.created" scheme="DCTERMS:W3CDTF" content="2001-10-04">
<meta name="THESIS.degree.name" content="Mgr.">
<meta name="THESIS.degree.level" content="magisterský">
<meta name="THESIS.degree.grantor" content="Univerzita Karlova v Praze. Filozofická fakulta. Ústav
informačních studií a knihovnictví">
<meta name="DC.publisher" content="KSV Jinonice">
<meta name="DC.identifier" scheme="DCTERMS:URI" content="http://doi.vskp.cz/120.12/31">
<meta name="DC.description" xml:lang="cs" content="Práce se zabývá problematikou veřejného
výpůjčního práva (dále VVP) jak v historickém kontextu, tak jeho současnou praxí v existujících
světových systémech. Práce je rozdělena do sedmi kapitol. -- První kapitola je věnována světové
terminologii užívané v oblasti VVP...">
</head>
...
```

2.3 Schémata kódování dat

Pro tvorbu a komunikaci metadat v národním i mezinárodním rámci jsou specifikovány potřebné standardy zápisu některých definovaných prvků (datum, jazyk, jméno apod.), které jsou závazné. Primárně se používají schémata DC (<http://purl.org/dc/elements/1.1/>) a DCTERMS (<http://purl.org/dc/terms/>).

2.3.1 Datum

Formát data se řídí dle standardu ISO 8601, jehož obsah je zahrnut i v dokumentu Konsorcia W3C (<http://www.w3.org/TR/NOTE-datetime>). Přípustné jsou různé varianty od zápisu jen roku přes zápis konkrétního dne (preferováno) až po přesnou definici času. Položky dne jsou odděleny pomlčkou, v případě času je povinný znak T a uvedení časové zóny (pro ČR „+01:00“)

Příklady:

```
<meta name="DC.date" content="2005" />
<meta name="DC.date" content="2005-01" />
<meta name="DC.date" content="2004-12-31" />
<meta name="DC.date" content="2004-12-31T19:20+01:00" />
```

2.3.2 Jazyk

Podle internetového standardu RFC 3066 (<http://www.ietf.org/rfc/rfc3066.txt>) je při kódování jazyka možné vycházet z normy ISO 639-1 (stanovuje dvoumístné kódování jazyka, za pomlčku lze uvést volitelně dvoumístný kód země) nebo normy ISO 639-2 (třímístné kódování jazyka, konkrétně ISO 639-2/B). V případě, že existuje dvoumístný kód, doporučuje se použít spíše kód dvoumístný než třímístný. Autorizovaný seznam kódů z ISO 639-2 (Codes for the representation of names of languages) je veden a udržován v rámci Registračního úřadu (authority) při LC (<http://www.loc.gov/standards/iso639-2/englangn.html>).

Neautorizovaný seznam kódů z normy ISO 639-2 je v překladu do češtiny z roku 2001 k dispozici také na URL: http://www.mpo.cz/xqw/webdav/-UTF8-/dms_mpo/getFileinternet/11534/8488/jazyky_en.pdf

Použitelná schémata:

- ISO 639-2
Kód jazyka složený ze tří znaků se specifikací země, kde je jazyk používán
- ISO 639-1
Kód jazyka složený ze dvou znaků

Příklad:

```
<meta name="DC.language" scheme="ISO 639-2" content="cze" />
<meta name="DC.language" scheme="ISO 639-1" content="cs" />
<meta name="DC.language" content="cs" />
<meta name="DC.title" xml:lang="en" content="Second title" />
```

2.3.3 Formát zdroje

Označení formátu elektronického souboru je prováděno většinou automaticky, ale jeho označení je významné při přenosu elektronického dokumentu, konverzi a při stahování dokumentu k uživateli. Standardem je běžně používaný Internet Media Types - MIME (Multipurpose Internet Mail Extensions) popsán v RFC 2046 (<http://www.ietf.org/rfc/rfc2046.txt>). Přehled typů MIME je k dispozici také na URL (je odkazován ze základních specifikací DCMI): (<http://www.iana.org/assignments/media-types/index.html>).

Příklad:

```
<meta name="DC.format" scheme="DCTERMS.IMT" content="application/pdf">
```

Další vybrané formáty dle MIME:

- HTML: text/html
- Adobe PDF: application/pdf
- Microsoft Word: application/msword

- Microsoft Excel: application/msexcel
- Microsoft RTF: application/rtf
- Obrázek JPEG: image/jpeg

2.3.4 Jména fyzických osob

Jména fyzických osob se zapisují ve formě: Příjmení, Křestní Jména

Příklad:

```
<meta name="DC.creator" content="Mácha, Karel Hynek">
```

2.4 Seznamy a číselníky

Seznamy/číselníky slouží jako jednotný zdroj používaných hodnot pro jednotlivé vybrané metadatové prvky. Doporučuje se používat mezinárodní a národní jmenné soubory autoritních záznamů a číselníky a dále zavést vlastní schéma EVSKP obsahující jednotlivé číselníky s trvalými odkazy na aktuální XML seznamy (lze využít při počítačovém zpracování – seznam nahrát a dát uživateli ve formuláři na výběr ve formě menu).

Všechny číselníky by měly být dostupné na unikátní perzistentní URL adrese. Budou mít doporučující charakter (tj. je možné použít i hodnotu neuvedenou v seznamu). Jednotlivé seznamy je možné dle požadavků jednotlivých uživatelů rozšiřovat o další hodnoty.

Pro e-VŠKP budou zapotřebí následující seznamy:

- Jmenné autority a rejstříky (autoři a osoby podílející se na VŠKP)
- Seznam slovníků předmětových selekčních jazyků
- Instituce přidávající hodnotu
- Studijní program a Studijní obor (dle akreditačních seznamů)
- Typy e-VŠKP (Zákon o VŠ/1998 (§ 45-47 a 72))
- Akademické tituly (Zákon o VŠ/1998 (§ 45-47 a 72))
- Instituce archivující a zpřístupňující VŠKP (Zákon o VŠ/1998 (§45-47 a 72))
- Jméno přidělované hodnotě (Zákon o VŠ/1998 (§ 45-47 a 72))
- Typ studijního programu (Zákon o VŠ/1998 (§ 45-47 a 72))
- Formát práce (RFC 2045 - MIME-IMT)
- Jazyk (RFC 3066)
- Práva k využívání

2.5 Příklad metadatového záznamu

Záznam reálné disertační práce s omezeným veřejně dostupným množstvím metadat (v kódu XHTML)

```
...
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1" />
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
<meta name="DC.title" xml:lang="cze" content="Nestandardní modely pro rozhodování a vyjednávání ekonomických subjektů týkající se ekonomických informací" />
<meta name="DC.creator" content="Hlaváček, Michal" />
<meta name="DC.description" xml:lang="cze" content="Tato disertační práce je věnována relativně nové oblasti ekonomie informací. Tento obor ekonomie je zajímavý mimo jiné tím, že se nachází na průniku zájmu jinak relativně striktně oddělených oblastí ekonomie jakými jsou například mikroekonomie, makroekonomie, teorie finančních trhů a finančního zprostředkování a pod. Disertační práce se skládá z pěti na sobě relativně nezávislými kapitolami, které mají charakter separátních článků, a které pokrývají výše uvedené oblasti. V první kapitole je prezentován model, který vysvětluje existenci finančního zprostředkování existencí informační asymetrie vycházející z obdobného modelu v článku Holmström, Tirole (1997). V kapitole druhé je prezentován mikroekonomický model pokoušející se modelovat předávání informací mezi dvěma subjekty, mezi kterými existuje pozitivní externalitní vazba, kde je modelována relativní ochota donátora předávat informaci v závislosti na jeho velikost. Ve třetí kapitole je
```

uveden model producenta informačních komodit (či komodit, které v sobě skrývají vysoké fixní výzkumné náklady), který má sice nulové mezní náklady a využívá možností monopolu, oproti tomu však čelí omezení souvisejícím s hrozbou vstupu konkurenta. Z jeho pohledu je tento vstup náhodnou událostí záviselí na výši zisku (čím vyšší zisk, tím vyšší pravděpodobnost vstupu konkurenta), kterou producent váží s opačným rizikem odchodu vlastníků z titulu nízkého zisku. Ve čtvrté kapitole je uvedena aplikace teorie transferu informací do makroekonomie, když je zde prezentován model difuze technologie se zpožděním mezi rozvinutou zemí vyvíjející nové technologie, a zemí zaostalou, která tyto technologie pouze se zpožděním kopíruje. Je ukázáno, že takovýto model může i při realistických předpokladech vysvětlit v realitě sledované rozdíly v ekonomické úrovni zemí (na rozdíl od standardních modelů s výzkumem a vývojem), přičemž zde klíčovou roli hraje právě faktor zpoždění. V kapitole jsou dále diskutovány faktory ovlivňující velikost faktoru zpoždění, především pak role lidského kapitálu. V závěrečné páté kapitole je prezentován probitový/logitový model odhadu přežívání podniků v letech 1998-2001. Na jeho základě je pak konstatováno, že v uvedené době hrála zřejmě difuze technologií v přežívání podniků významnou roli." />

<meta name="DC.publisher" content="Univerzita Karlova. Fakulta sociálních věd. Institut ekonomických studií" />

<meta name="DC.contributor.advisor" content="Cahlík, Tomáš" />

<meta name="DC.date.created" scheme="DCTERMS.W3CDTF" content="2004-08-01" />

<meta name="DC.type" content="text.dissertation" />

<meta name="DC.format" scheme="DCTERMS.IMT" content="text/pdf" />

<meta name="DC.identifier" scheme="URL"

content="http://ies.fsv.cuni.cz/storage/work/406_hlavacek.pdf" />

<meta name="DC.language" scheme="ISO639-2" content="cze" />

<meta name="thesis.degree.name" content="Ph.D." />

<meta name="thesis.degree.level" content="doktorský" />

<meta name="thesis.degree.discipline" content="Ekonomie/Ekonomické teorie" />

<meta name="thesis.degree.grantor" content="Univerzita Karlova. Fakulta sociálních věd" />

...

3 Příspěvatelé

Text návrhu standardu EVSKP-MS byl jako interní dokument připravován postupně Pracovní skupinou (viz zpracovatelé) Odborné komise pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací AKVŠ, a to zhruba v období od června 2004 do listopadu 2005. Návrh standardu v jeho první veřejné verzi (EVSKP-MS, verze 0.1) byl vícenásobně diskutován na pravidelných zasedáních Odborné komise. Předkládaná verze byla v základních částech schválena dne 3. 10. 2005. K formulacím textu v rámci diskusí přispěli zejména:

- Alena Balvínová, Univerzita Karlova v Praze, Přírodovědecká fakulta, Středisko vědeckých informací
- Eva Bratková, Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví
- Lukáš Budínský, Univerzita Tomáše Bati ve Zlíně, Ústřední knihovna
- Eva Bulínová, Univerzita Karlova v Praze, Ústav výpočetní techniky
- Jiřina Čechová, Technická univerzita v Liberci, Univerzitní knihovna
- Simona Esteřáková, Univerzita Tomáše Bati ve Zlíně, Ústřední knihovna
- Iva Horová, Knihovna Akademie múzických umění Praha
- Věra Jurmanová-Volemanová, Masarykova univerzita v Brně, Filozofická fakulta, Ústřední knihovna
- Šárka Kostecká, Ostravská univerzita v Ostravě, Univerzitní knihovna
- Martin Krčál, Masarykova univerzita v Brně, Fakulta sociálních studií, Ústřední knihovna
- Adam Kretschmer, Technická univerzita v Liberci, Ústřední knihovna
- Jarmila Krkošková, Ostravská univerzita v Ostravě, Univerzitní knihovna
- Jaroslav Lipovský, Veterinární a farmaceutická univerzita v Brně, Ústřední knihovna
- Jan Mach, Vysoká škola ekonomická v Praze, Centrum informačních a knihovnických služeb
- Martina Nedomová, Masarykova univerzita v Brně, Pedagogická fakulta, Ústřední knihovna
- Gabriela Opletalová, Vysoké učení technické v Brně, Fakulta stavební, Knihovnické a informační centrum
- Jiří Šimonek, Ostravská univerzita v Ostravě, Centrum informačních technologií
- Radka Tichá, Západočeská univerzita v Plzni, Univerzitní knihovna
- Miloš Trávníček, Vysoké učení technické v Brně, CVIS, Oddělení databází
- Milada Vondrová, Vysoké učení technické v Brně, Areálová knihovna Purkyňova